Some Recent Advances in Algebraic Multigrid

Parallel AMG and AMGe (element)

Robert D. Falgout

Van Emden Henson

Jim Jones

Ulrike Meier Yang

Center for Applied Scientific Computing

Lawrence Livermore National Laboratory

October 7, 1998

Parallel AMG requires parallel algorithms for these steps:

- *The Setup Phase*
  - Coarse Grid Selection
  - Construction of Prolongation operator, P
  - Construction of coarse-grid operators by Galerkin method, $P^T A P$

- *The Solve Phase*
  - Residual Calculation
  - Relaxation
  - Prolongation
  - Restriction

Parallelizing the *Solve* Phase

- *The Solve Phase*
  - **Residual Calculation**
    - **entails Matvec:** $y \leftarrow \alpha A x + \beta y$
  - **Relaxation**
    - **We use Jacobi rather than Gaus-Seidel**
    - **entails scaled Matvec-like operation**

$$x \leftarrow x + D^{-1}(b - Ax)$$

Parallelizing the *Solve* Phase, II

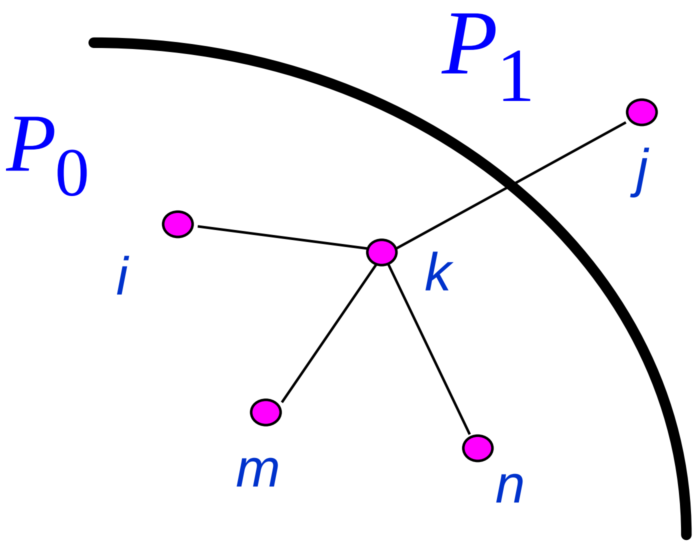
- *The Solve Phase*

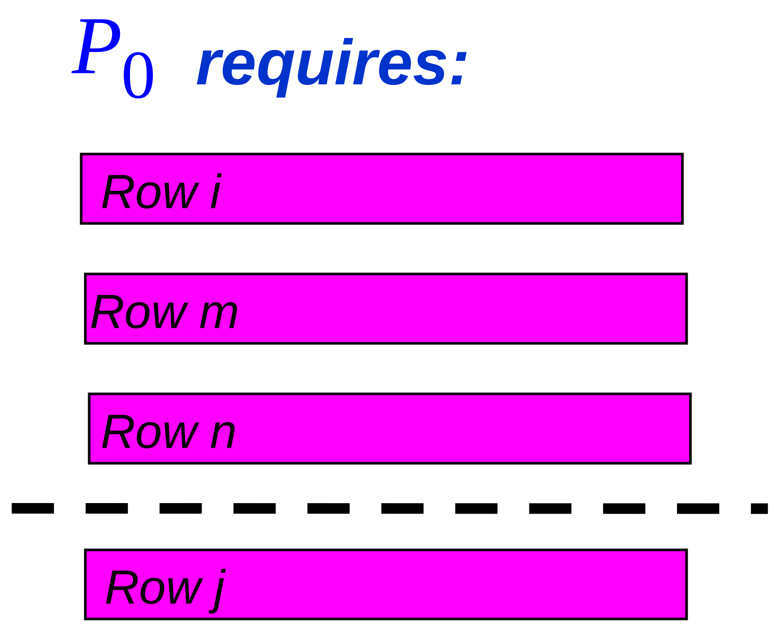
  — **Prolongation**
    - **Requires a simple Matvec, but on a rectangular matrix. May not be available in some common toolkits, but readily built and easily parallelizable**

  — **Restriction**
    - **Requires a MatvecT, the product of the transpose of a rectangular matrix. Not generally available in toolkits, but is easily constructed.**

The Parallel *Setup* phase

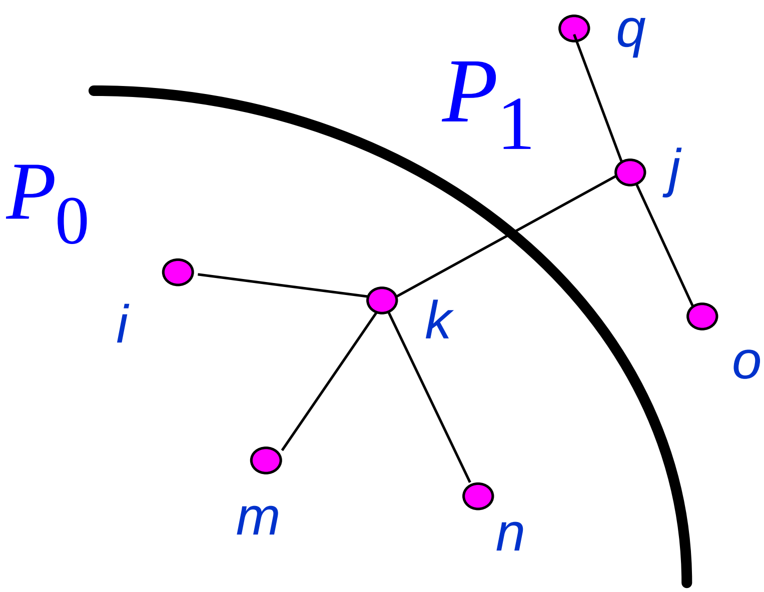
- **Construction of Prolongation operator, P, requires "processor boundary" equations (ghost point information), and can be accomplished using toolkit functions.**

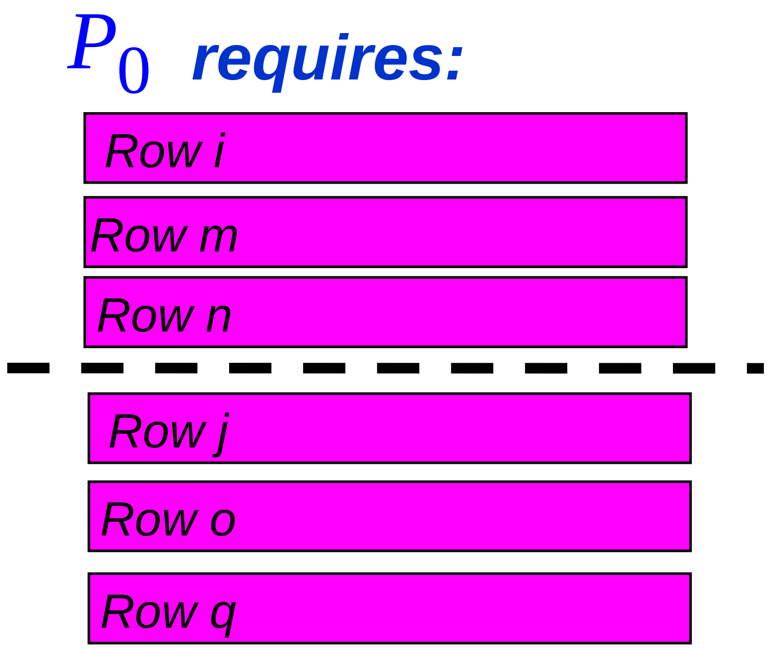




The Parallel ***Setup*** phase

- **Construction of coarse-grid operators by Galerkin method, $P^T A P$, requires two layers of processor boundary data.**





The Parallel *Setup* phase, II

- **Selection of the coarse-grid points**
  - **The *main challenge* of parallelizing AMG. The standard algorithm is inherently serial, requiring pathlength-two updates after each C-point is selected before work can continue.**
  - **We have developed a parallel coarsening algorithm that uses a Luby-Jones-Plassman like MIS algorithm to select coarse-grid points based on an “influence measure” that favors points that influence many other points.**

Useful definitions

- The variable u_i **depends** on the variable u_j if the jth coefficient in the ith equation is large compared to the other off-diagonal coefficients in the ith equation. That is, if

$$-a_{ij} > \theta \max_{j \neq i} (-a_{ij})$$

(assumes M-matrix)

- If j **depends** on k, then k **influences** j, which is denoted graphically by:

$$j \bullet \longrightarrow \bullet\, k.$$

- The set of coarse-grid variables is denoted C.
- The set of coarse-grid variables used to interpolate the value of the fine-grid variable u_i is denoted C_i.

S, the matrix of influence and dependence

- **Define S to be the adjacency matrix of the graph of influence associated with the operator, A.**

$$S_{ij} = \begin{cases} 1 & i \Rightarrow j \\ 0 & else \end{cases}$$

(*i* depends on *j*
j influences *i*)

- **The nonzero columns in the *i*th row of S, denoted $S_{i:}$ form the set of dependencies of the point *i*.**
- **The nonzero rows in the *i*th column of S, denoted $S_{:i}$ form the set of influences of the point i.**

Standard AMG:
Choosing the Coarse Grid

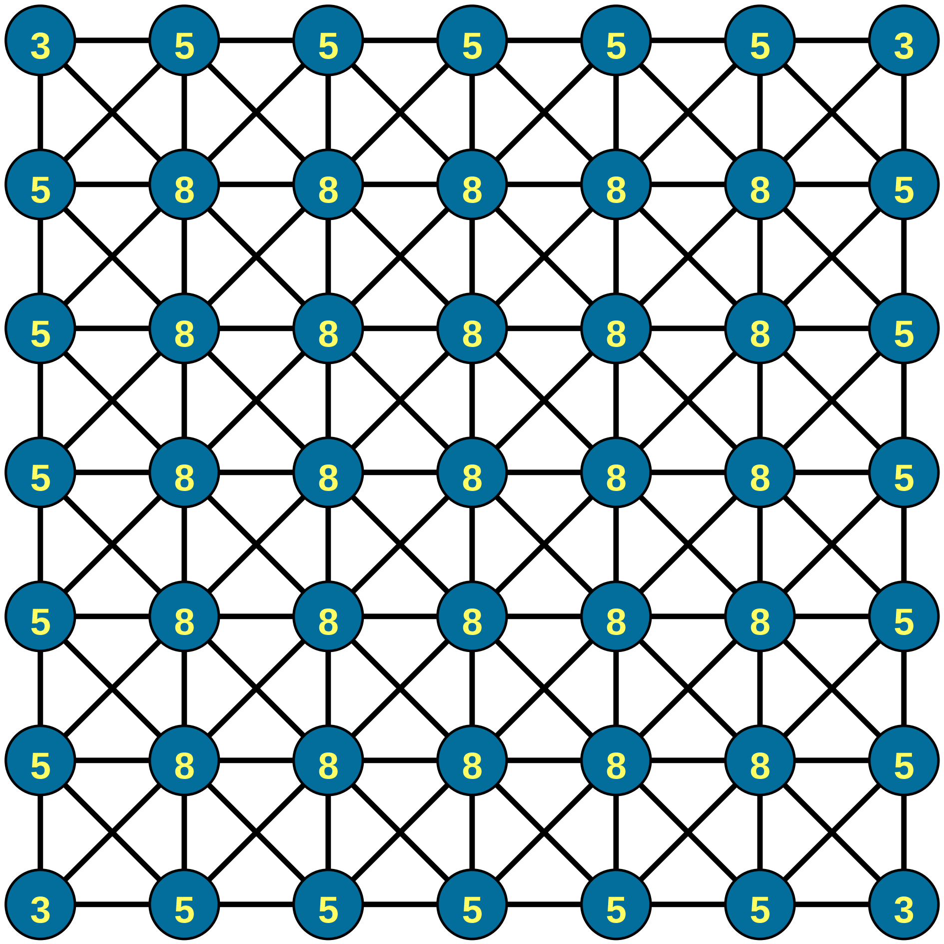
- **Two Criteria**

  — **(C1) For each fine-grid point *i*, each $j \in S_i$: should either be in *C* or should be dependent on at least one point in C_i .**

  — **(C2) *C* should be a maximal subset with the property that no C-point influences another C-point.**

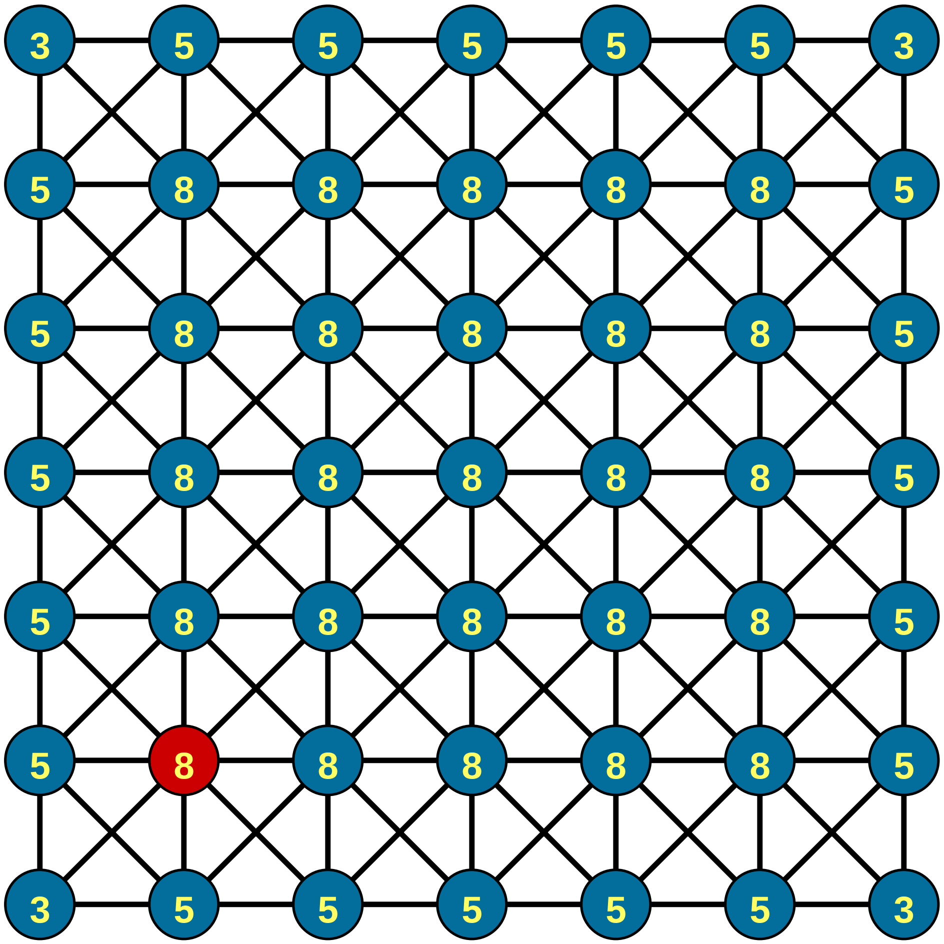
- **Satisfying both (C1) and (C2) is sometimes impossible. We use (C2) as a guide while enforcing (C1).**

Standard AMG coarsening: algorithm is inherently sequential



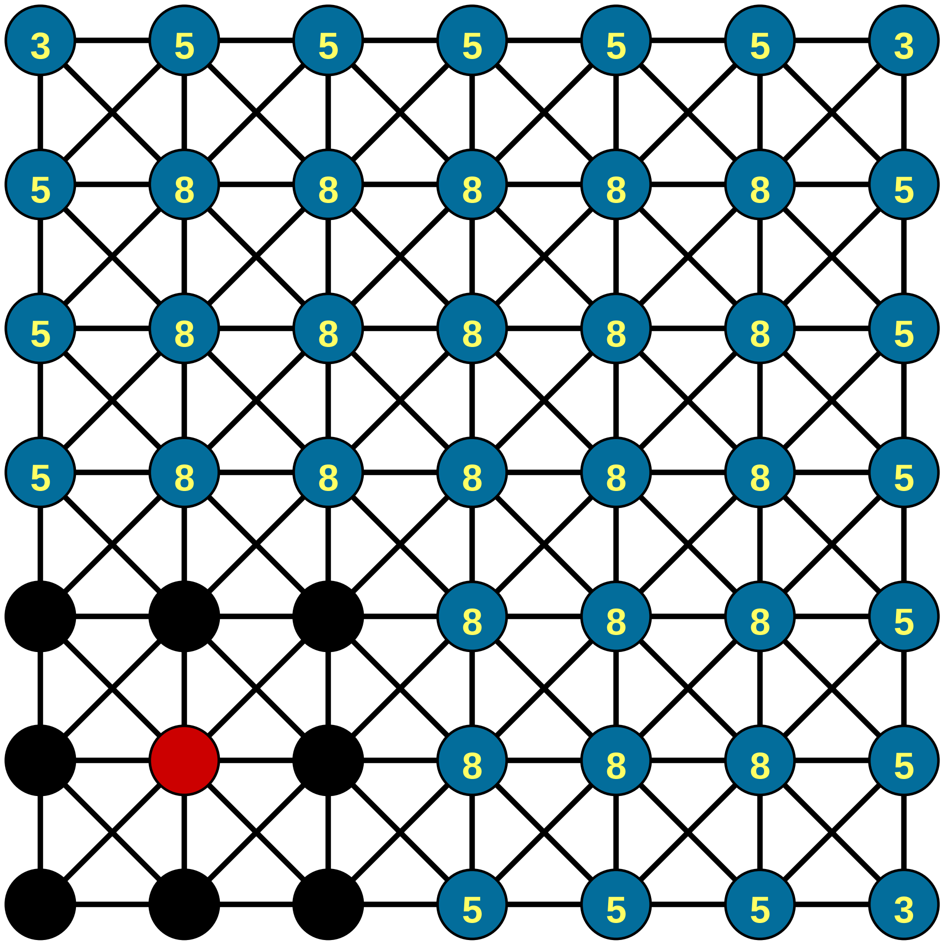
- **select C-pt with maximal measure**
- **select neighbors as F-pts**
- **update measures of F-pt neighbors**

Standard AMG coarsening: algorithm is inherently sequential



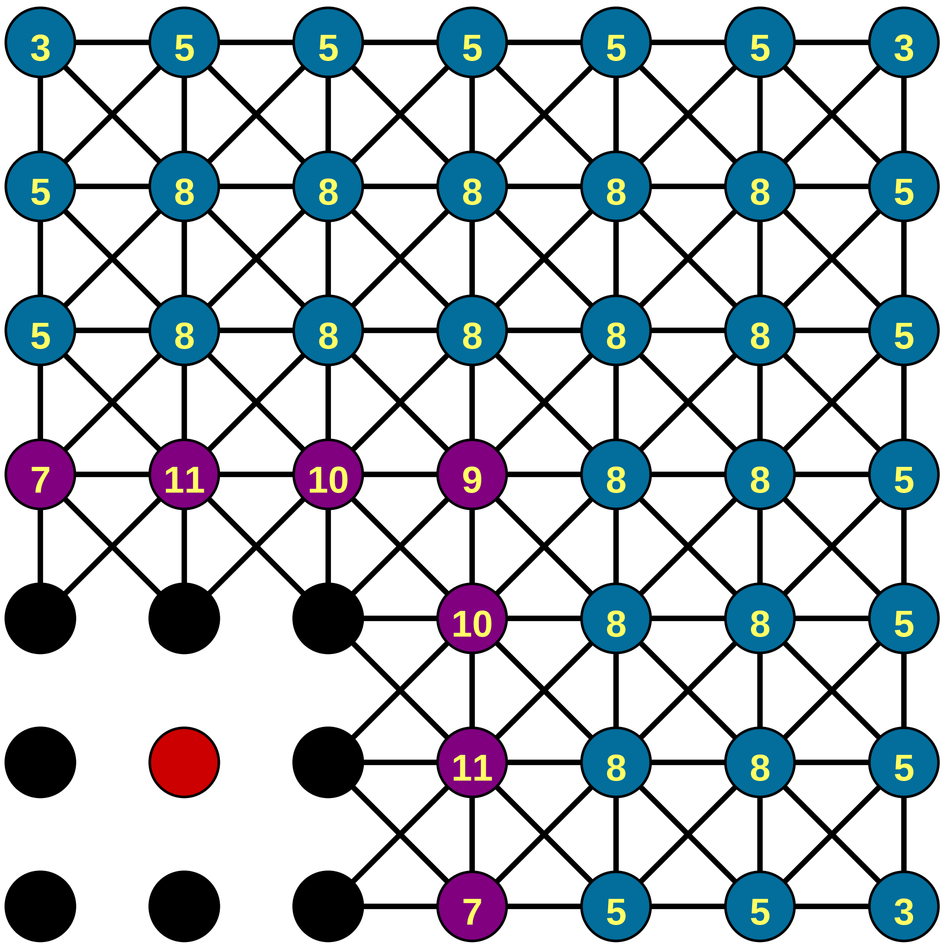
- select C-pt with maximal measure
- select neighbors as F-pts
- update measures of F-pt neighbors

Standard AMG coarsening: algorithm is inherently sequential



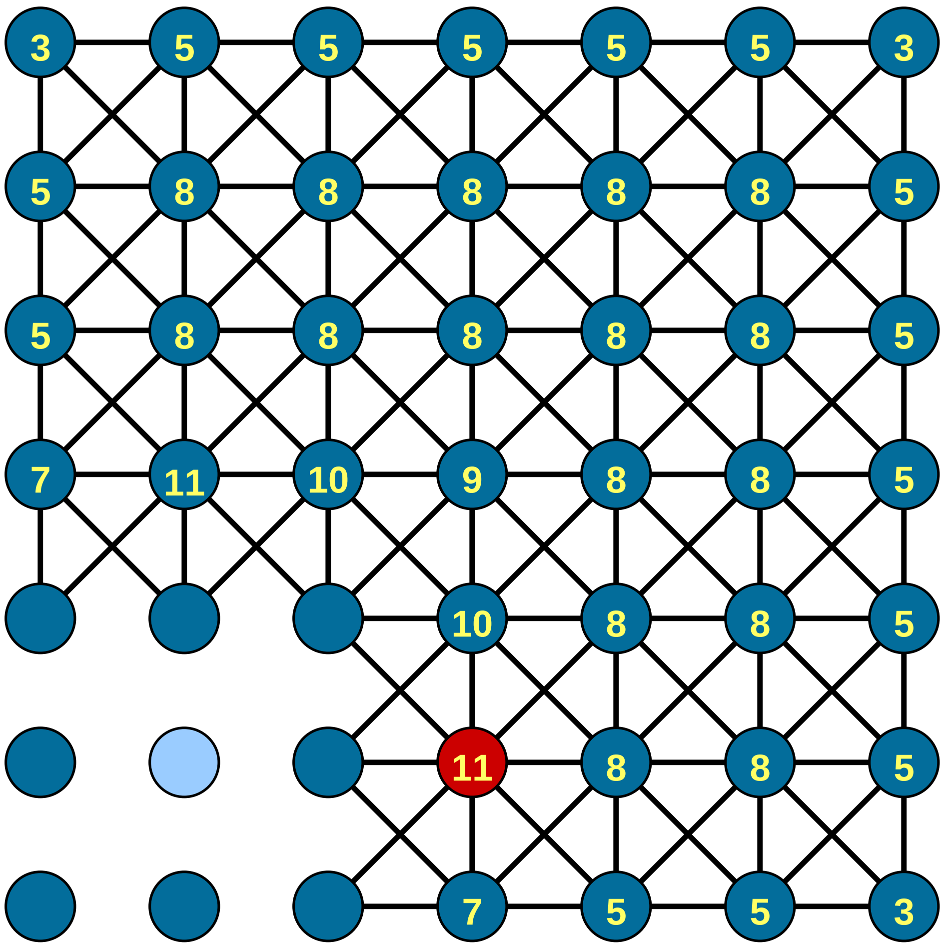
- select C-pt with maximal measure
- select neighbors as F-pts
- update measures of F-pt neighbors

Standard AMG coarsening: algorithm is inherently sequential



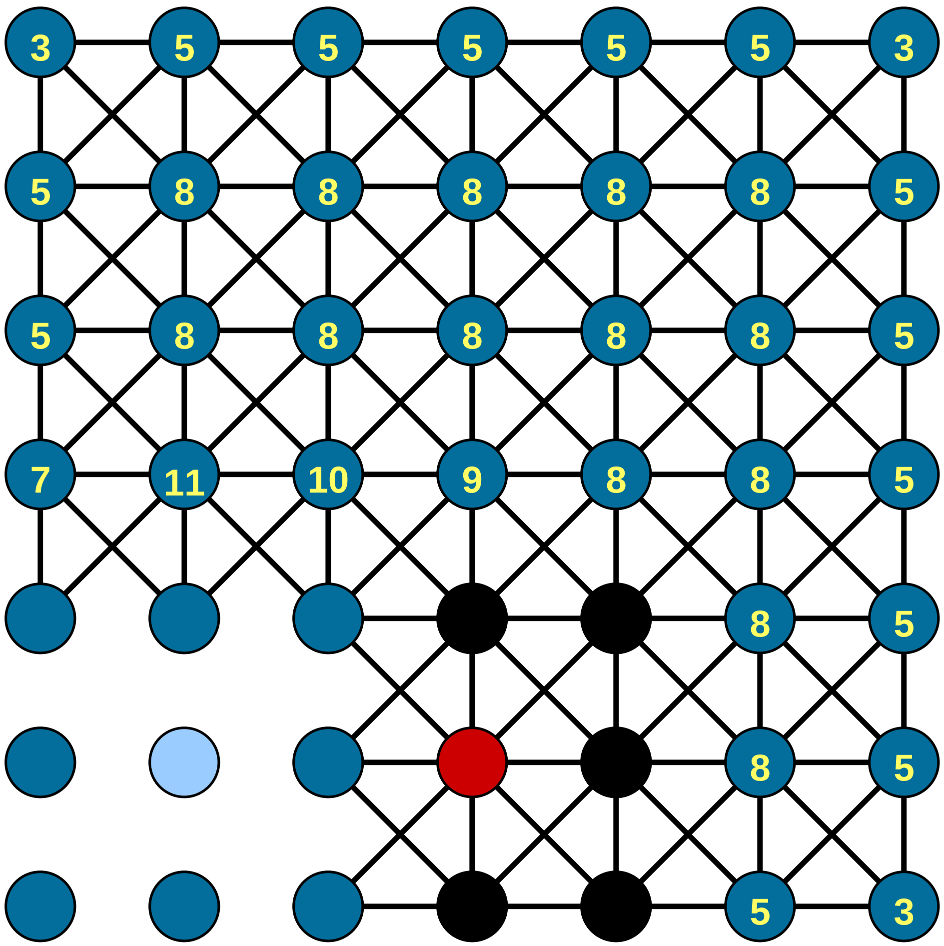
- select C-pt with maximal measure
- select neighbors as F-pts
- update measures of F-pt neighbors

Standard AMG coarsening: algorithm is inherently sequential



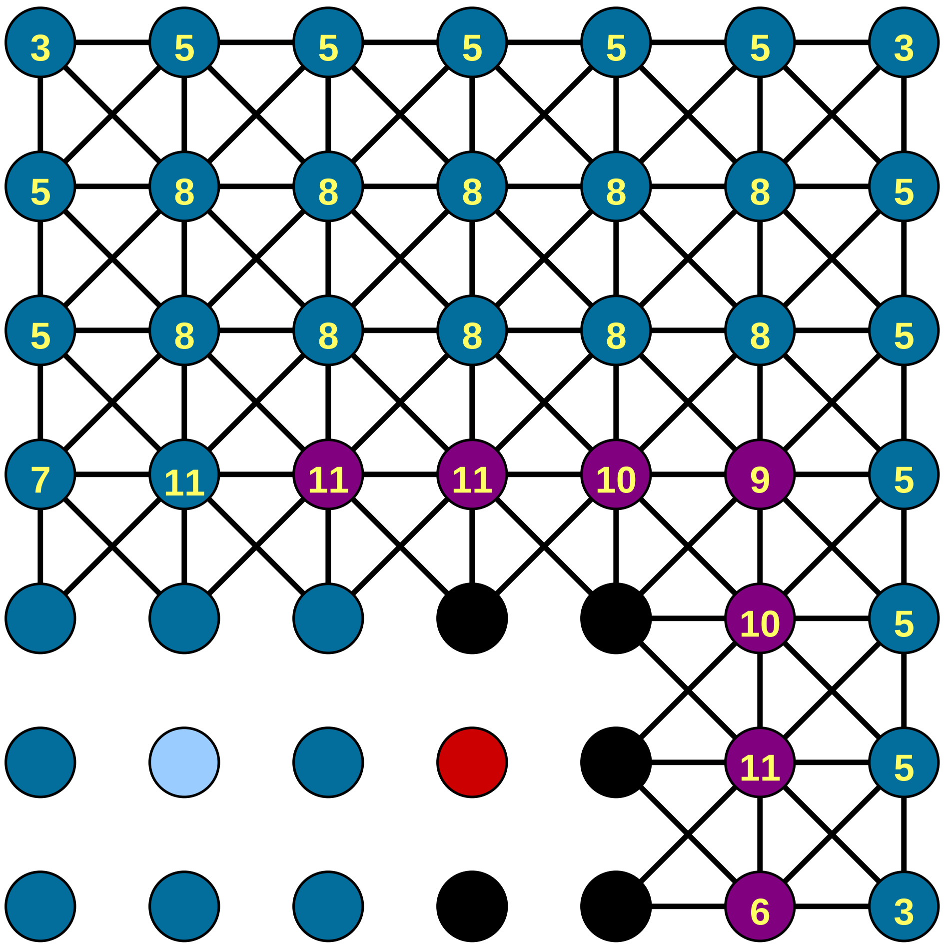
- select C-pt with maximal measure
- select neighbors as F-pts
- update measures of F-pt neighbors

Standard AMG coarsening: algorithm is inherently sequential



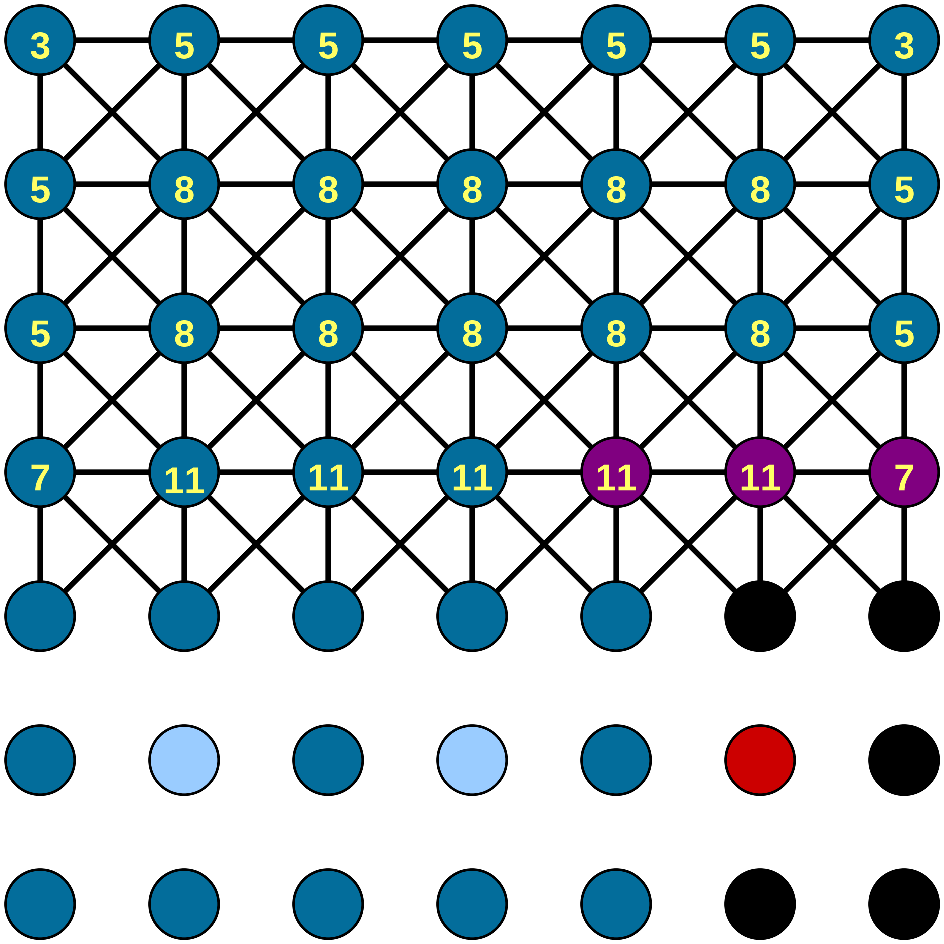
- select C-pt with maximal measure
- select neighbors as F-pts
- update measures of F-pt neighbors

Standard AMG coarsening: algorithm is inherently sequential



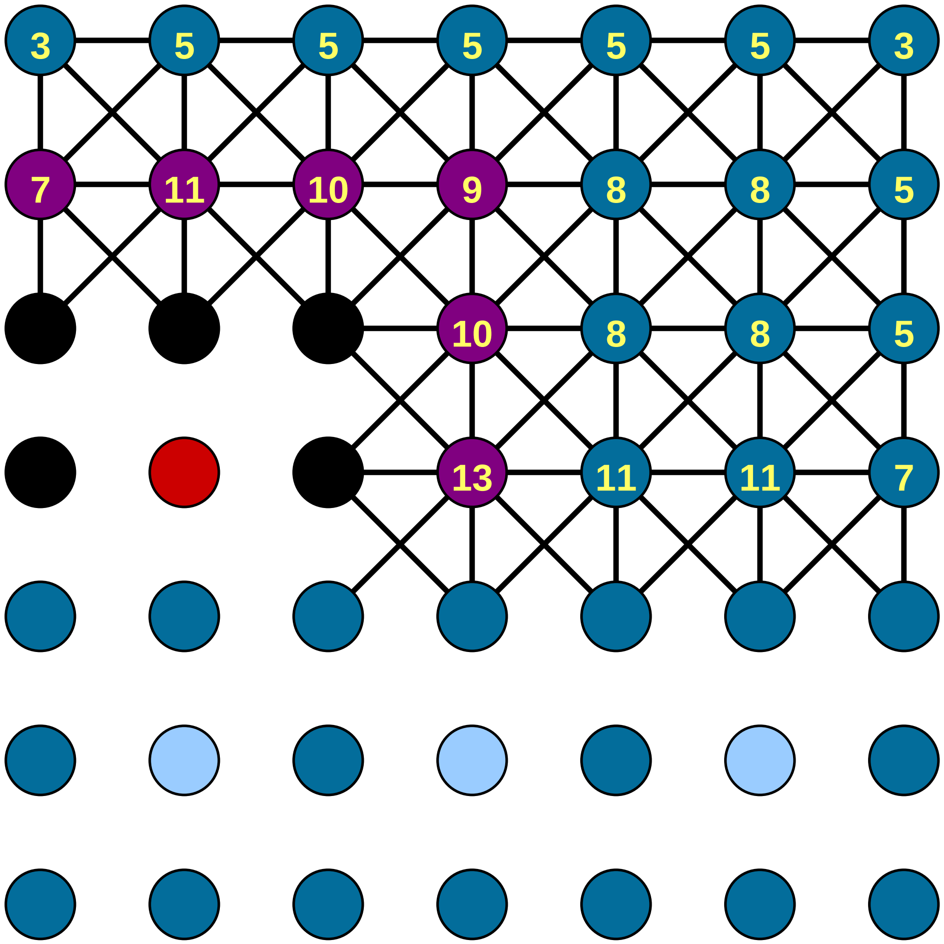
- select C-pt with maximal measure
- select neighbors as F-pts
- update measures of F-pt neighbors

Standard AMG coarsening: algorithm is inherently sequential



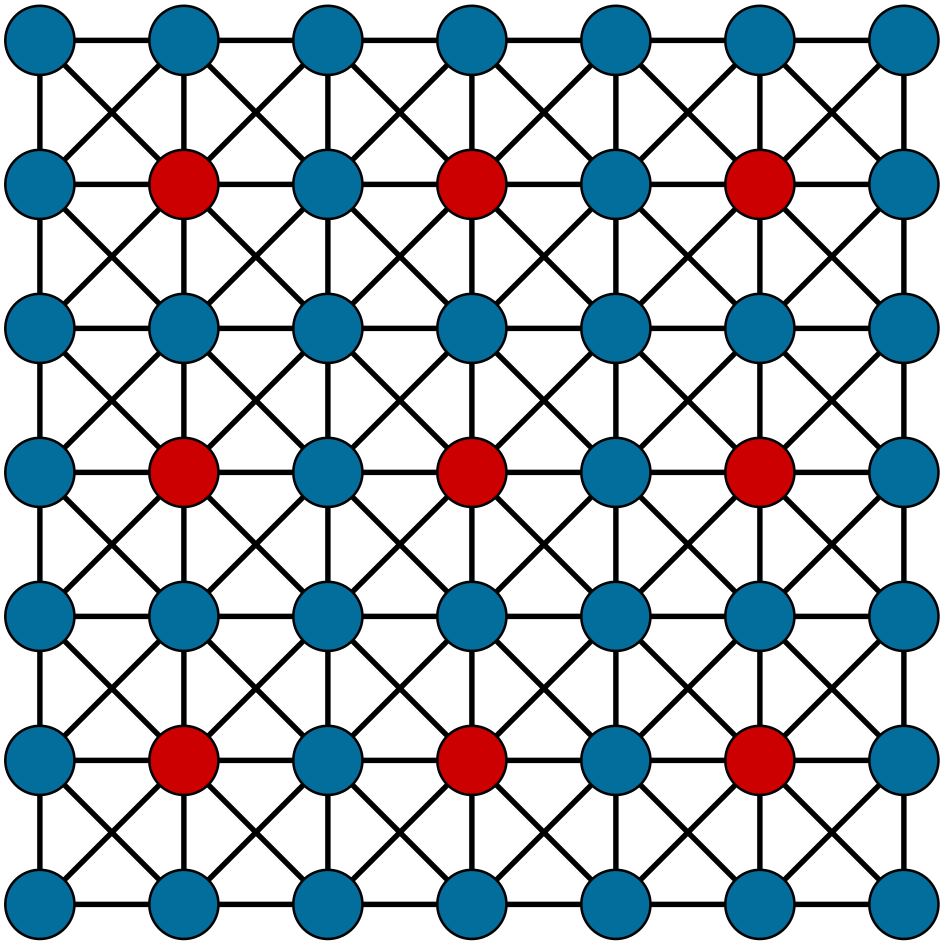
- select C-pt with maximal measure
- select neighbors as F-pts
- update measures of F-pt neighbors

Standard AMG coarsening: algorithm is inherently sequential



- select C-pt with maximal measure
- select neighbors as F-pts
- update measures of F-pt neighbors

Standard AMG coarsening: algorithm is inherently sequential



- select C-pt with maximal measure
- select neighbors as F-pts
- update measures of F-pt neighbors

ParAMG Coarsening

- Create a "measure" of each point, consisting of the number of influences of the point, plus a random number in [0,1].
- Select a set of points whose measure exceeds that of all points they influence or depend on. Set is independent by construction, may be maximal, (needn't be). This can readily be done in parallel, since once the random values are distributed, the only action is read-only!
- Perform ParAMG heuristics (described below) on the set of points selected above. Can be done in distributed fashion, requiring a synchronization at the end of the step when the set is exhausted.

ParAMG Coarsening Heuristic 1: the effect of coarse points

- **The values at a C-point is not interpolated, hence C-points won't need to interpolate from neighbors they depend on. Those neighbors have lessened "value" as potential C-points themselves.**

  — **For each neighbor, j, that influences c:**
    - **subtract 1 from measure[j]; and**
    - **remove the edge S_{cj} from the graph**

ParAMG Heuristic 2: neighbors dependent on a common C-point

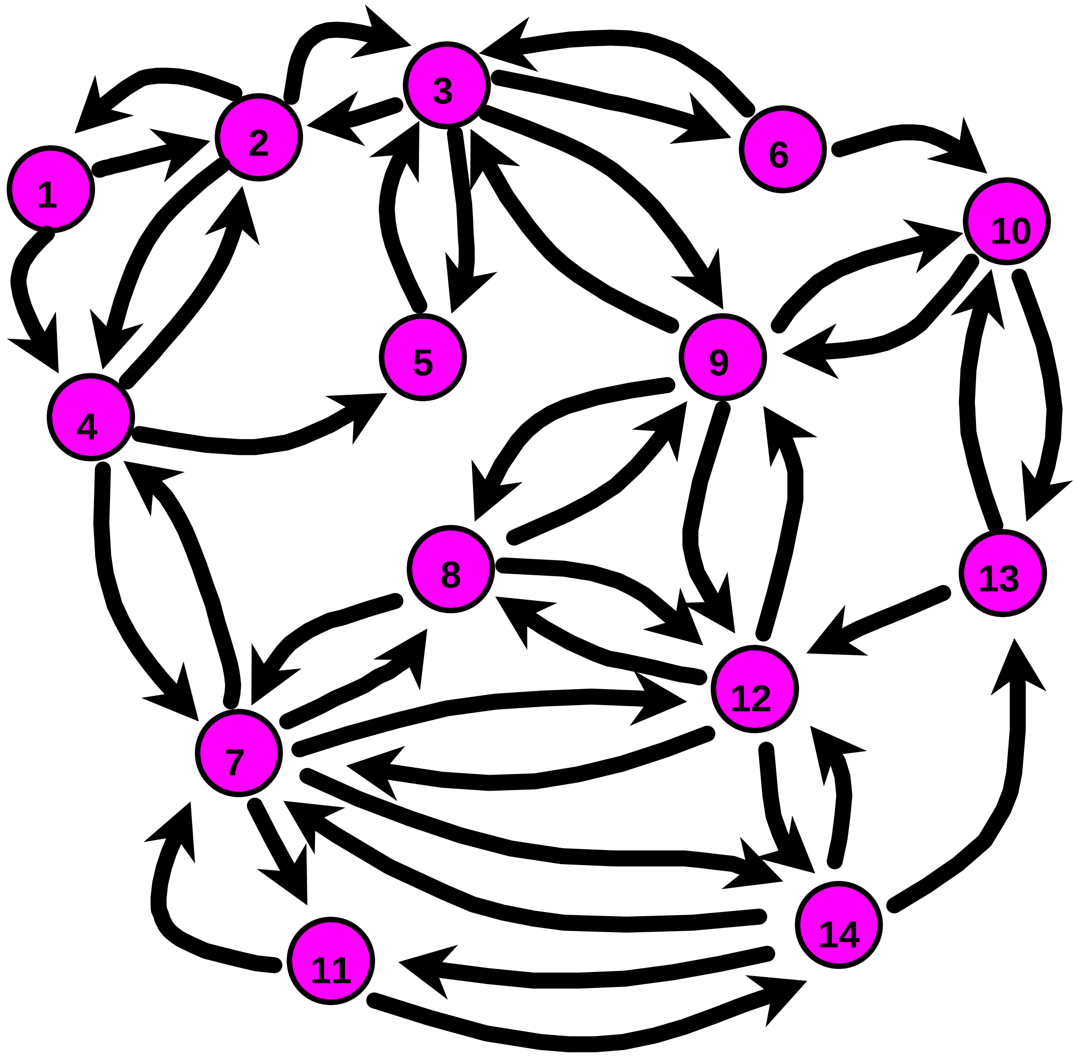
- **If k and j both depend on a given C-point, and j influences k, then the value of j as a coarse point is lessened, since k can be interpolated from C.**

  — **For each j that C influences: (i.e., $S_{jC} \neq 0$)**
    - **delete S_{jC}**
    - **for each k that j influences:**
      - **if k depends on C:**
        - **subtract 1 from measure[j];**
        - **remove edge S_{kj}**

ParAMG Coarsening

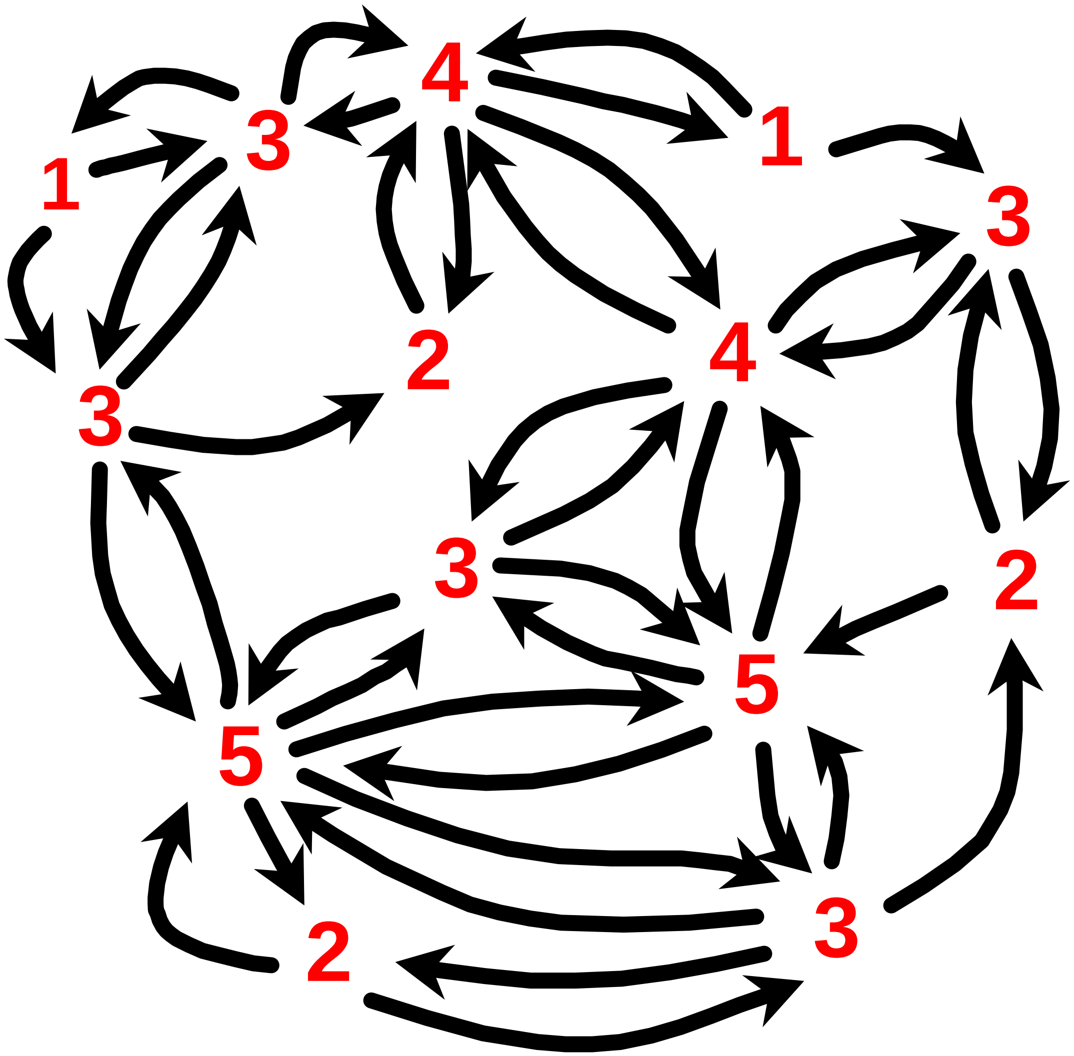
- Repeat the process on those vertices and edges remaining in the graph. A vertex is removed when all its edges are removed.
- The process continues until all points have either been selected as a C-point by the independent set picker or have been removed from the graph by virtue of measure[j] --> 0.
- The union independent sets is the coarse grid. All other points form the fine grid.

Selecting the coarse-grid points



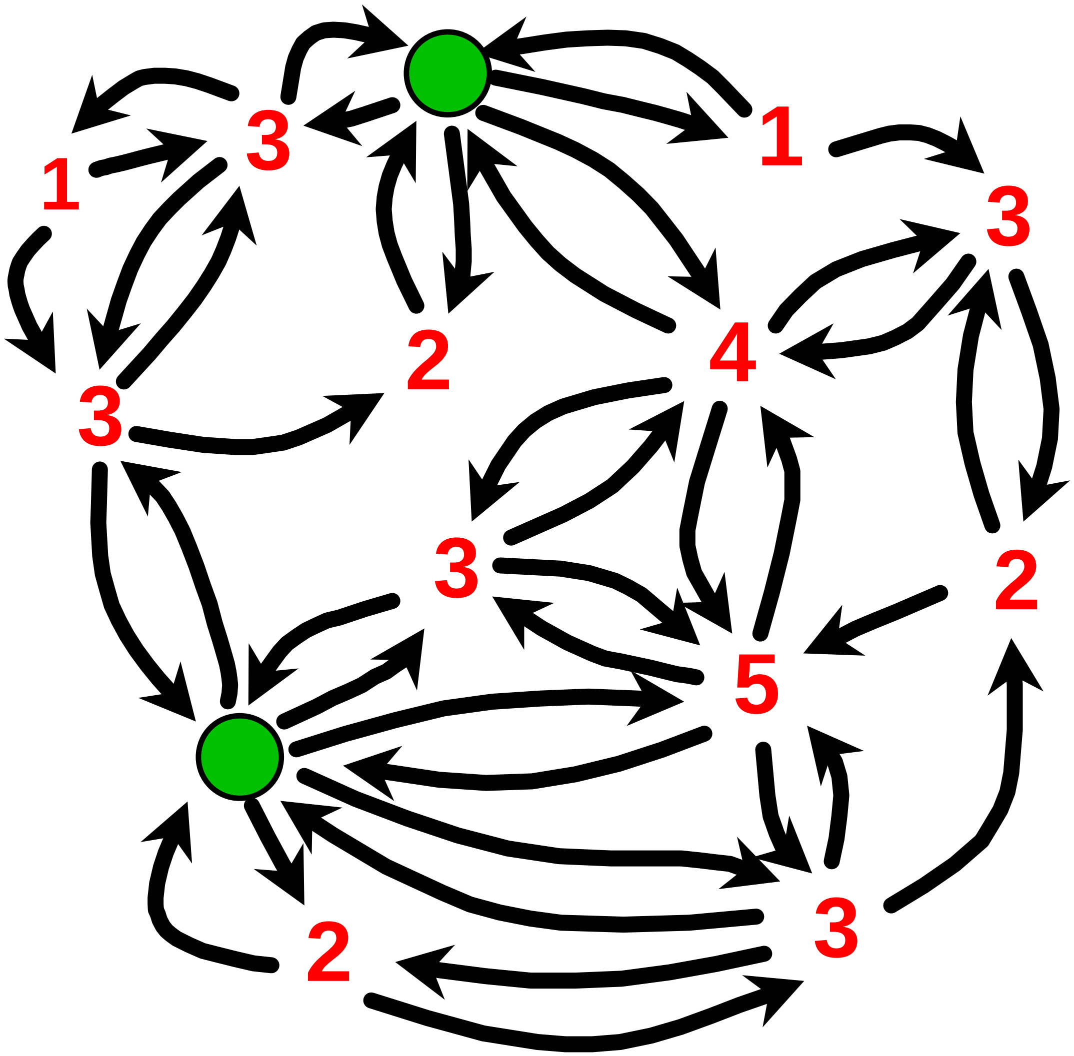
$$
\begin{bmatrix}
 & * & & * & & & & & & & & & & \\
* & & * & * & & & & & & & & & & \\
 & * & & & * & * & & & * & & & & & \\
 & * & & & * & & * & & & & & & & \\
 & & * & & & & & & & & & & & \\
 & & * & & & & & & & * & & & & \\
 & & & * & & & & * & & & * & * & & * \\
 & & & & & & * & & * & & & * & & \\
 & & * & & & & & * & & * & & * & & \\
 & & & & & & & & * & & & & * & \\
 & & & & & & * & & & & & & & * \\
 & & & & & & * & * & * & & & & & * \\
 & & & & & & & & & * & & * & & \\
 & & & & & & * & & & & * & * & * &
\end{bmatrix}
$$

Selecting the coarse-grid points



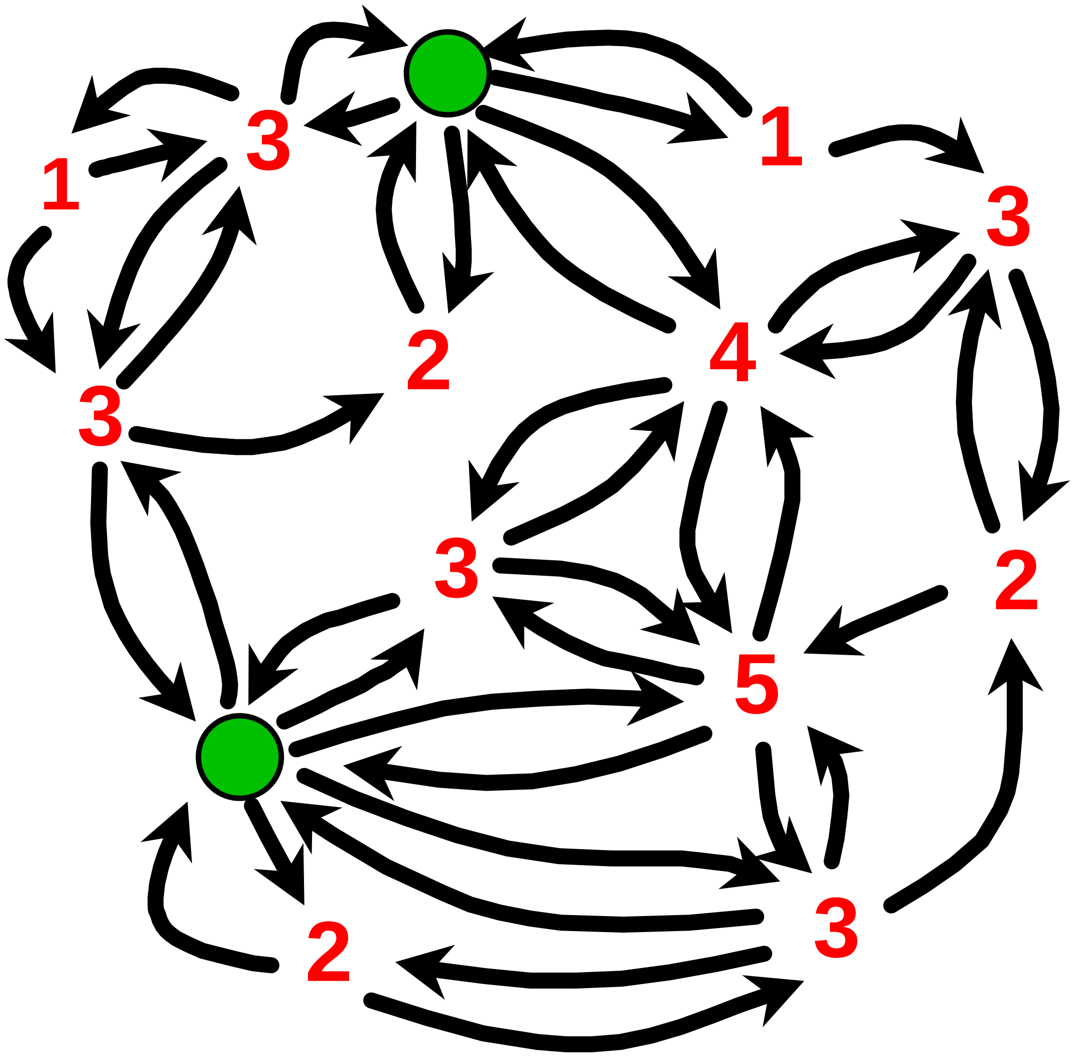
Determine the "measure" for each point: the number of other points influenced (arrows pointing into the point) plus a random number between 0 and 1 (not shown).

Selecting the coarse-grid points



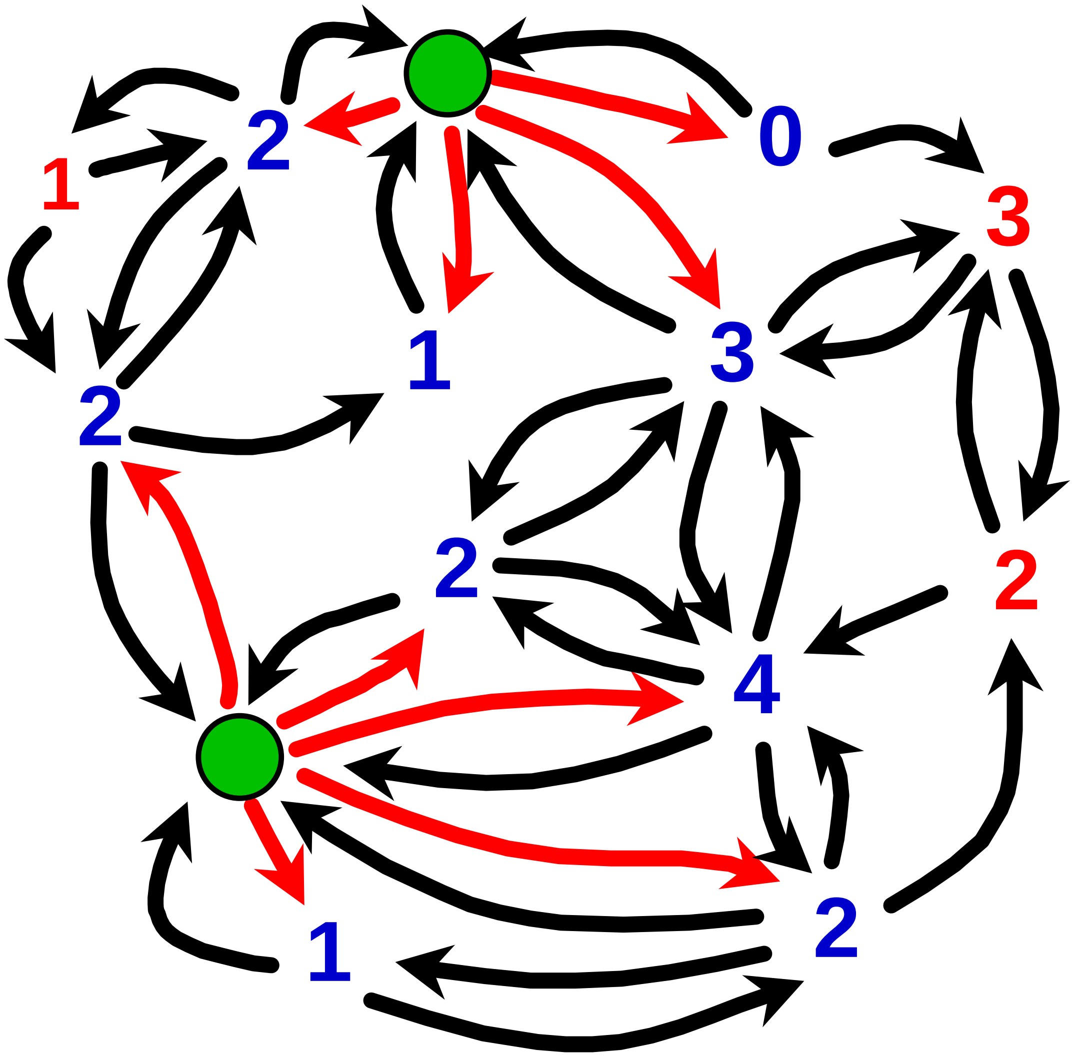
Choose as C-points an independent set of points whose measures are greater than those of all their neighbors.

Selecting the coarse-grid points



C-points will not be interpolated.

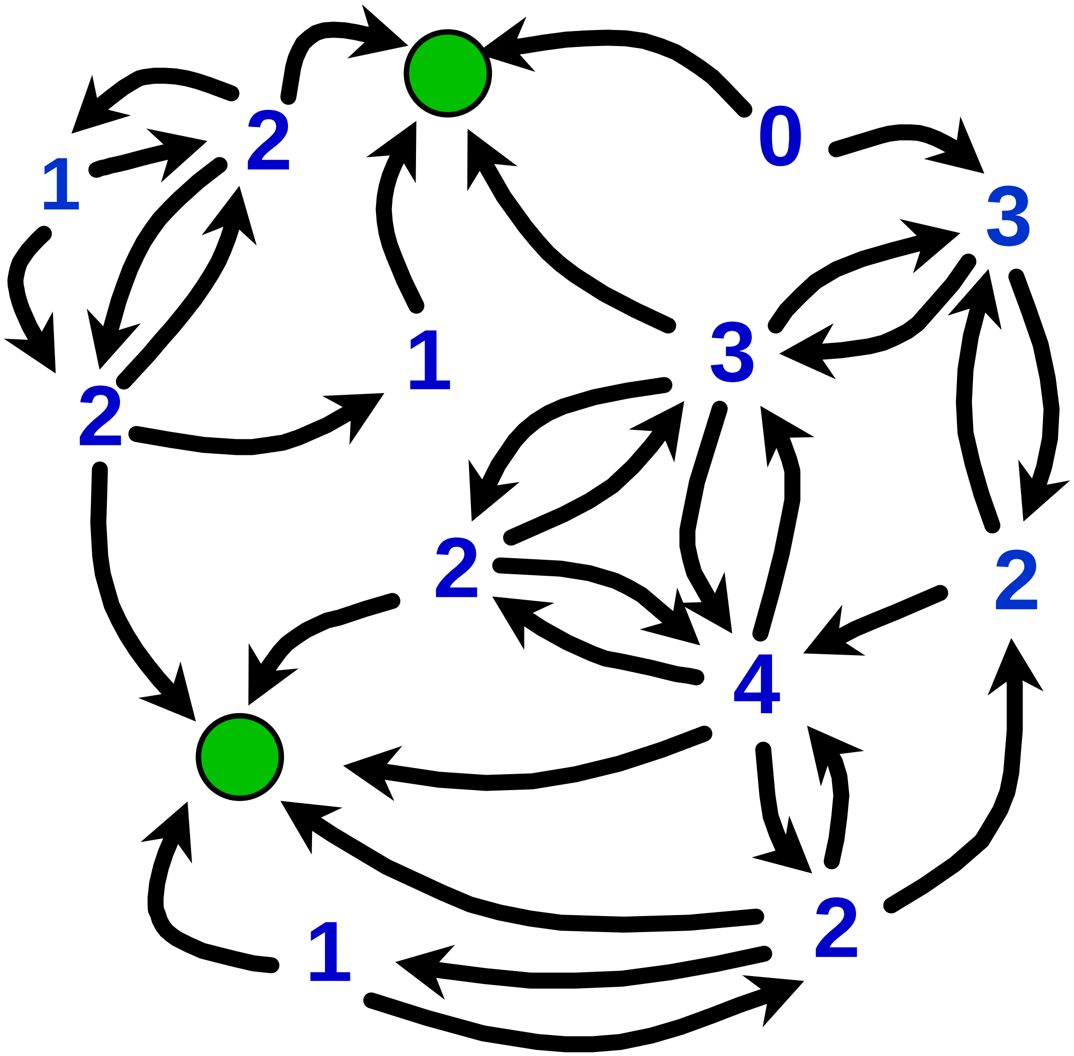
Selecting the coarse-grid points



C-points will not be interpolated.

Lower the measures of the points that influence these C-points.

Selecting the coarse-grid points

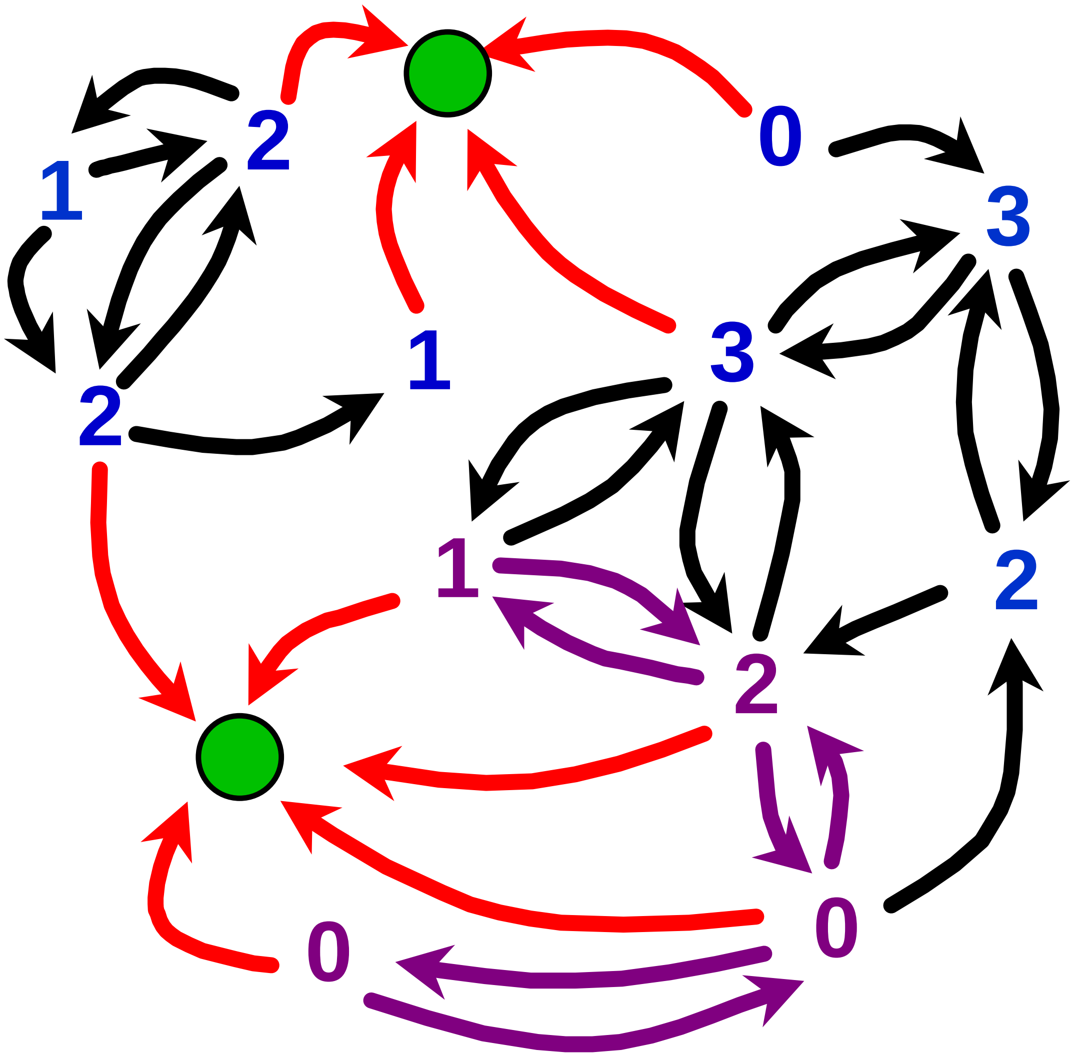


C-points will not be interpolated.

Lower the measures of the points that influence these C-points.

Remove the edges showing this influence from the graph.

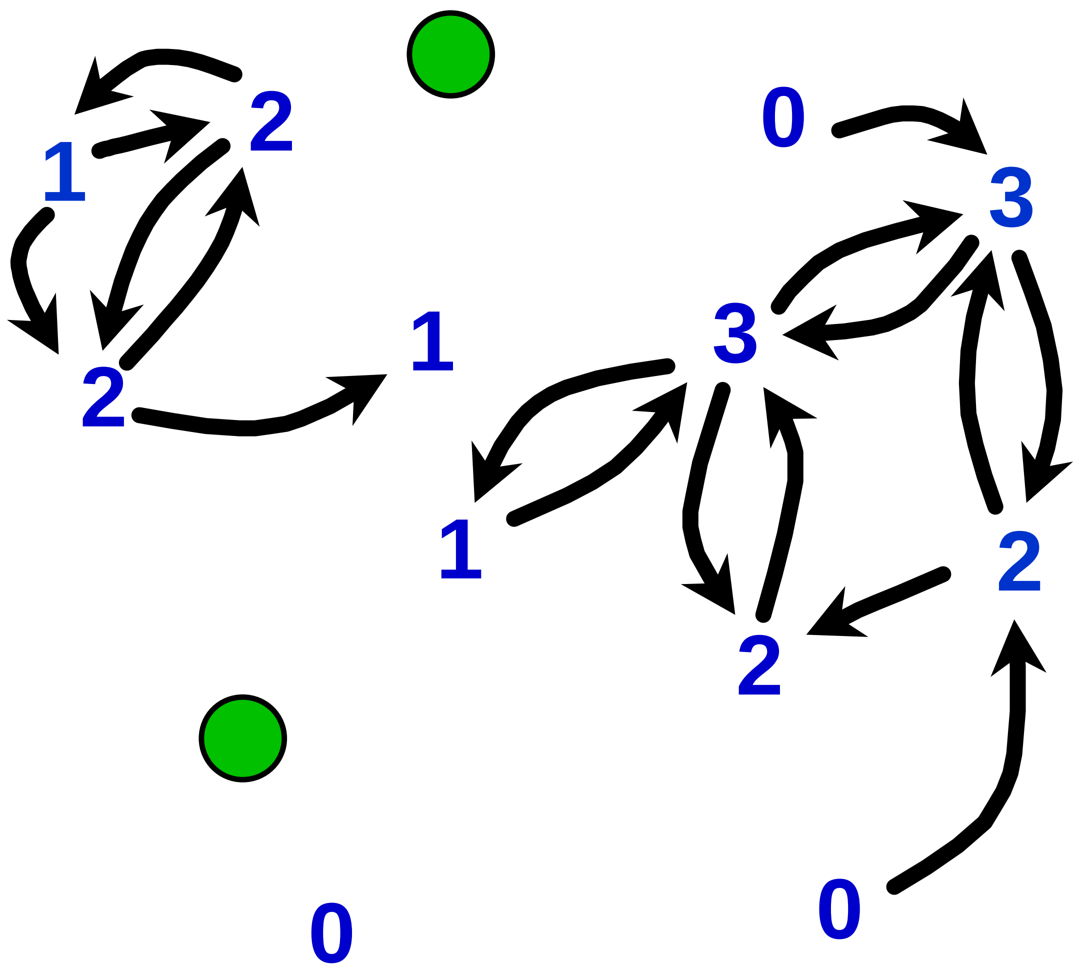
Selecting the coarse-grid points



F-points influenced by a common C-point don't interpolate each other:

Lower the measure of each point P, influenced by C, for every other point P influences that also depends on C

Selecting the coarse-grid points

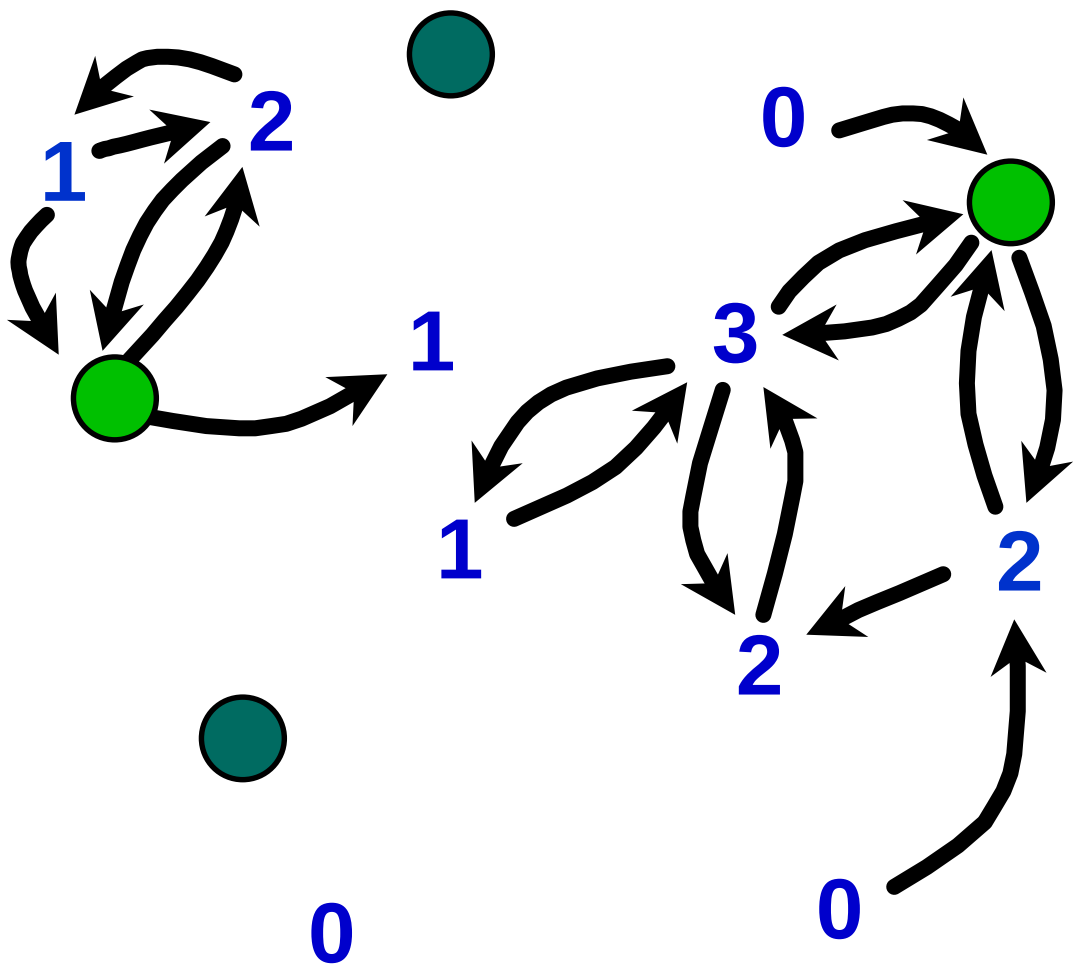


F-points influenced by a common C-point don't interpolate each other:

Lower the measure of each point P, influenced by C, for every other point P influences that also depends on C

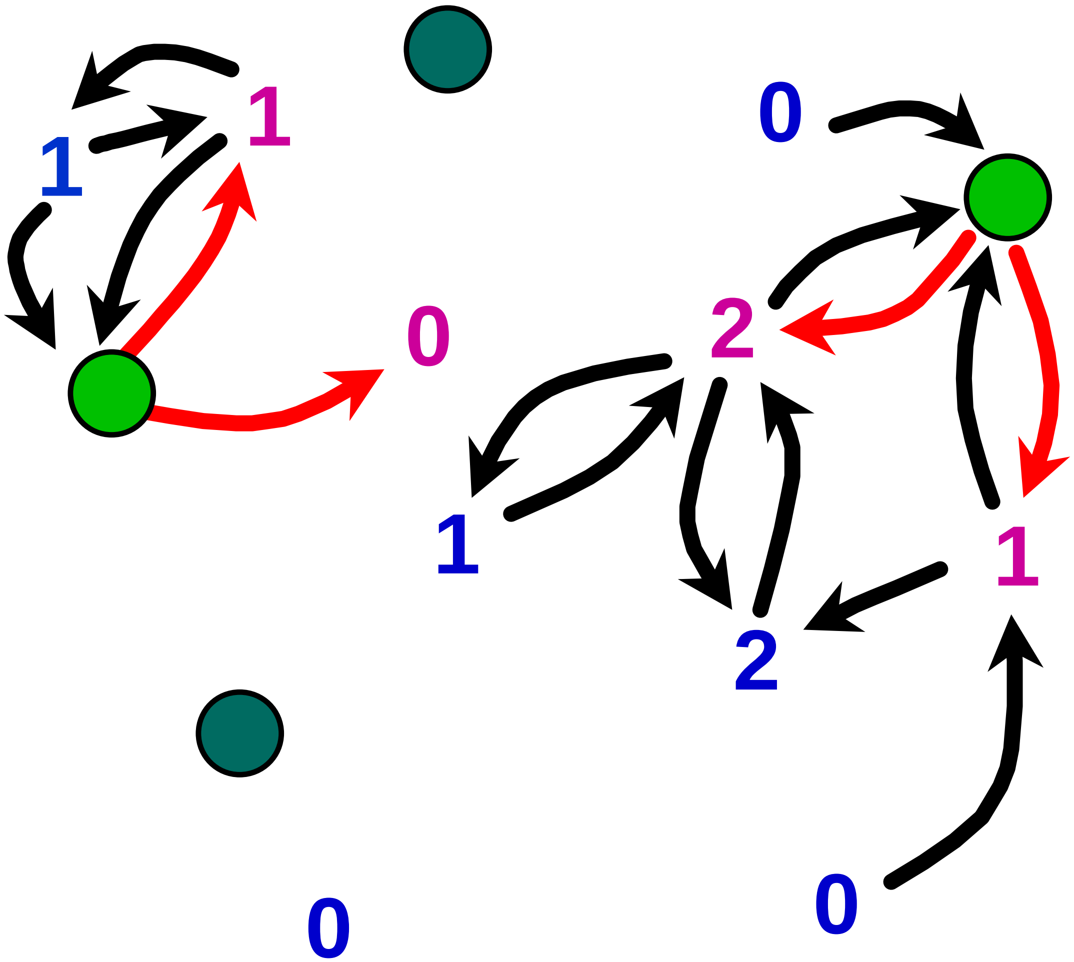
Remove the edges showing this influence from the graph.

Selecting the coarse-grid points



From the graph that remains Choose as C-points an independent set of points whose measures are greater than those of all their neighbors.

Selecting the coarse-grid points

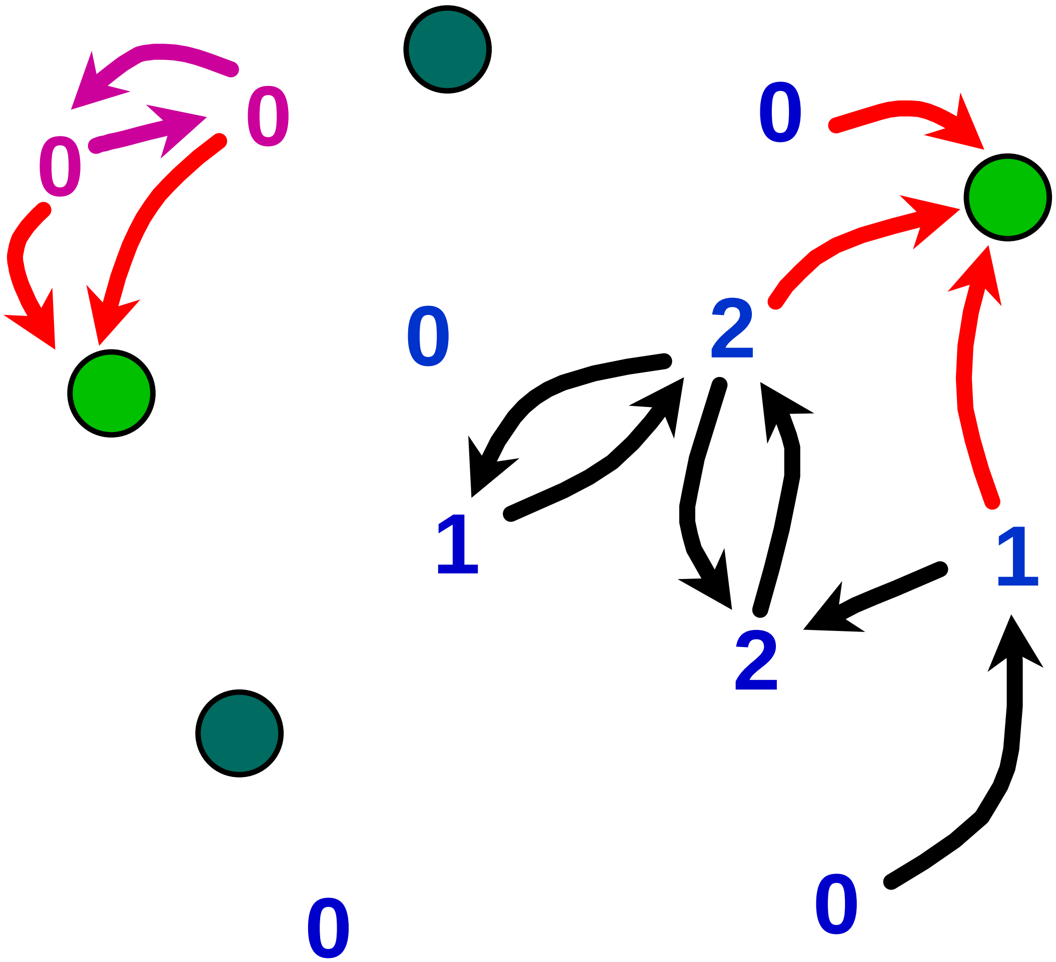


C-points will not be interpolated.

Lower the measures of the points that influence these C-points.

(Next, remove the edges showing this influence from the graph.)

Selecting the coarse-grid points

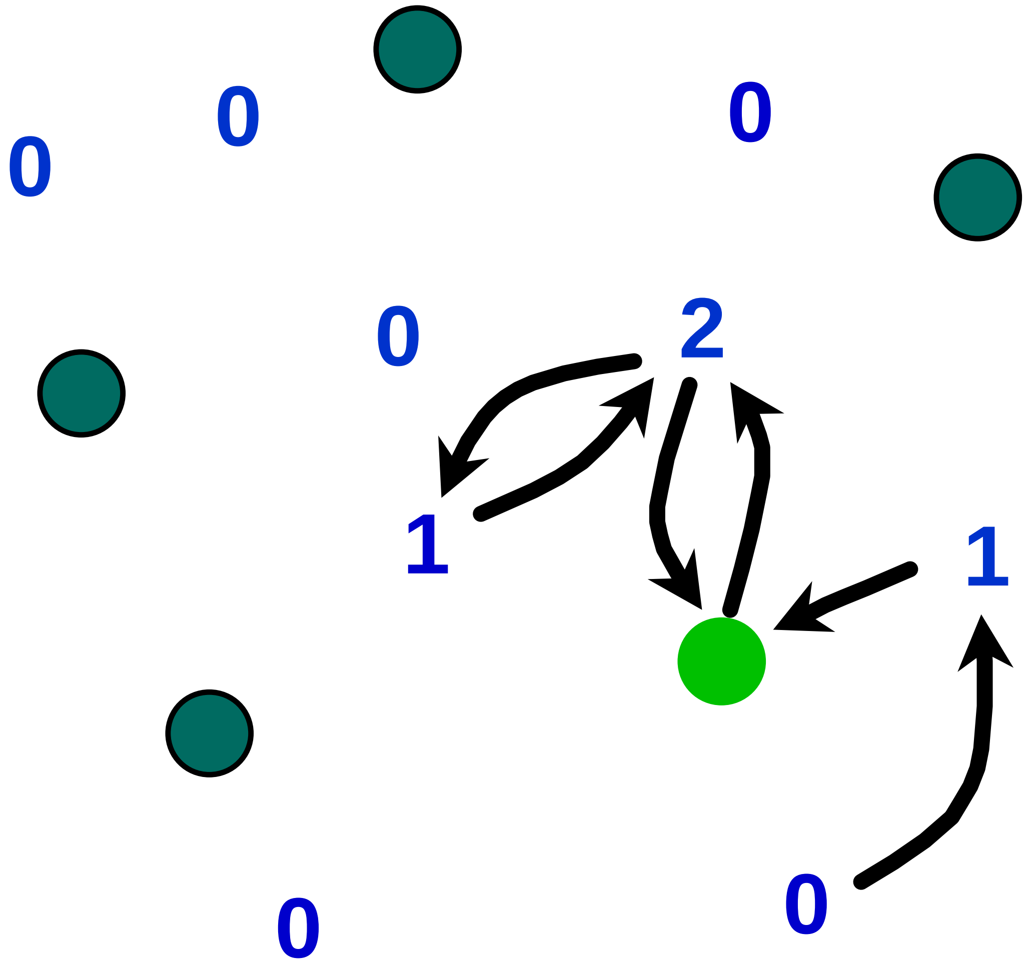


F-points influenced by a common C-point don't interpolate each other:

Lower the measure of each point P, influenced by C, for every other point P influences that also depends on C

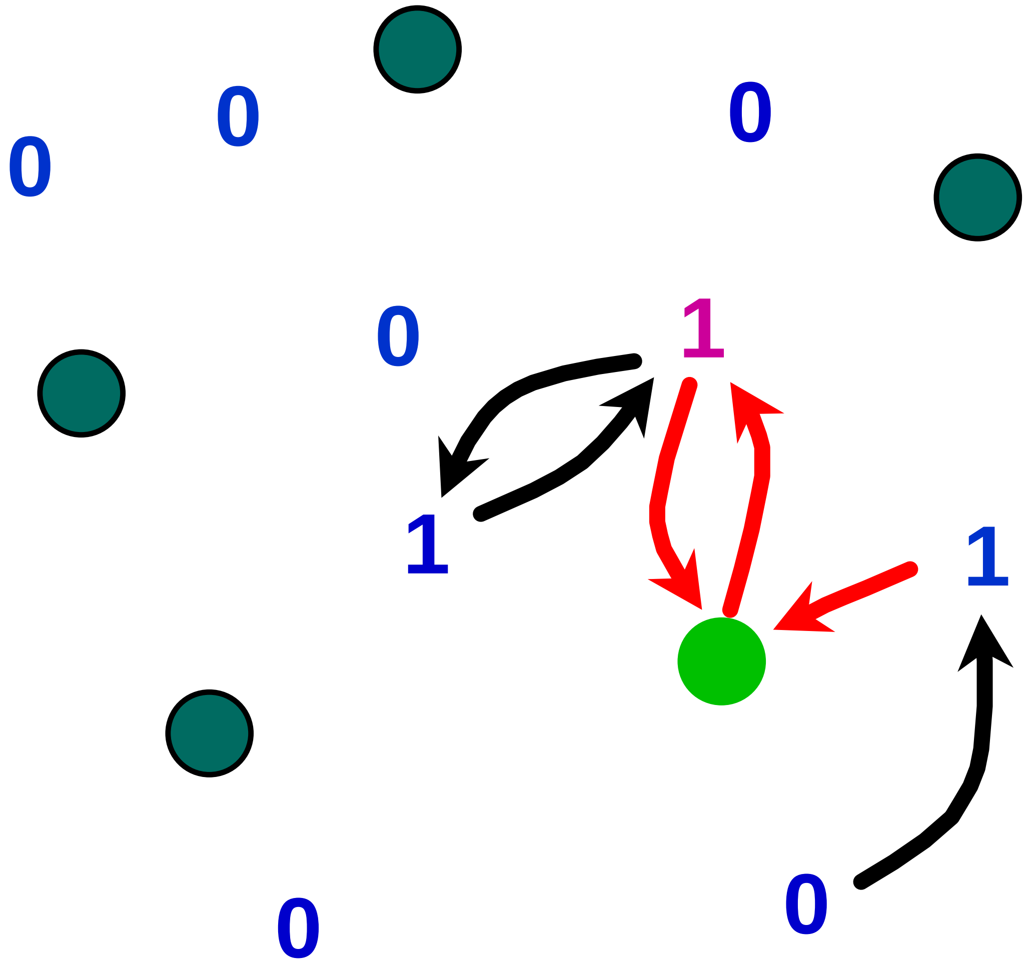
(Next, remove the edges showing this influence from the graph.)

Selecting the coarse-grid points



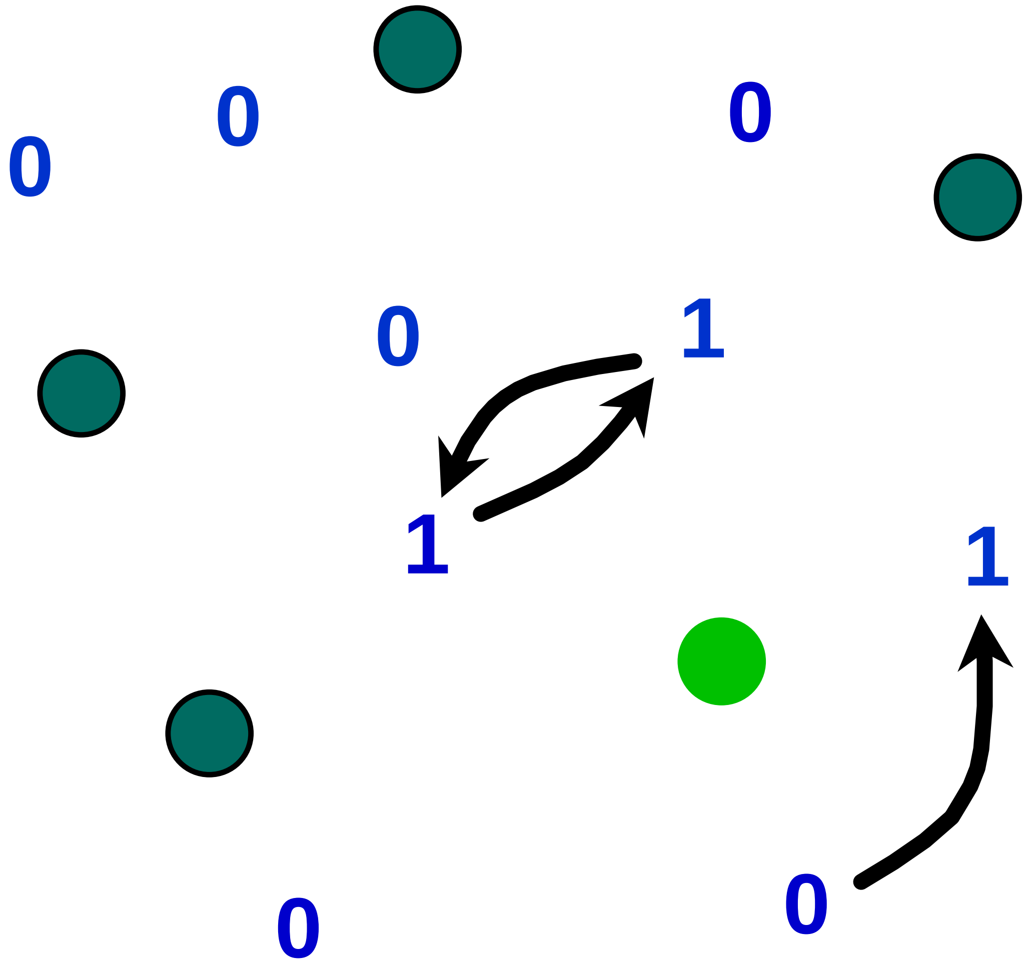
From the graph that remains Choose as C-points an independent set of points whose measures are greater than those of all their neighbors.

Selecting the coarse-grid points



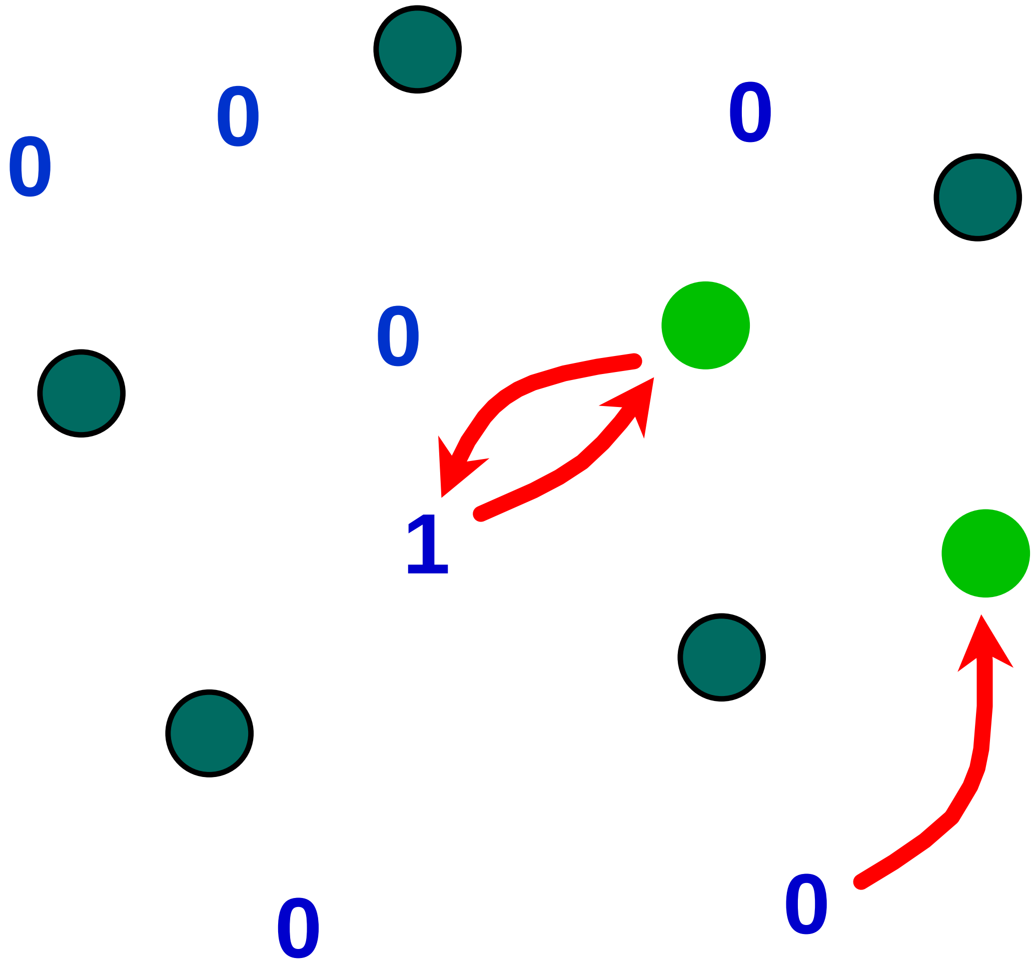
Perform the Heuristics.

Selecting the coarse-grid points



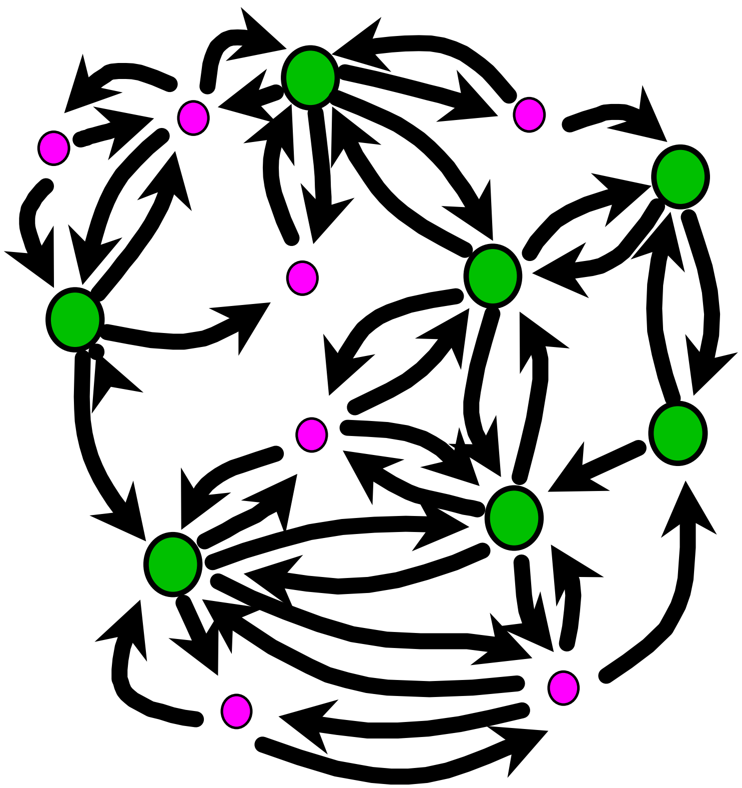
Perform the Heuristics.

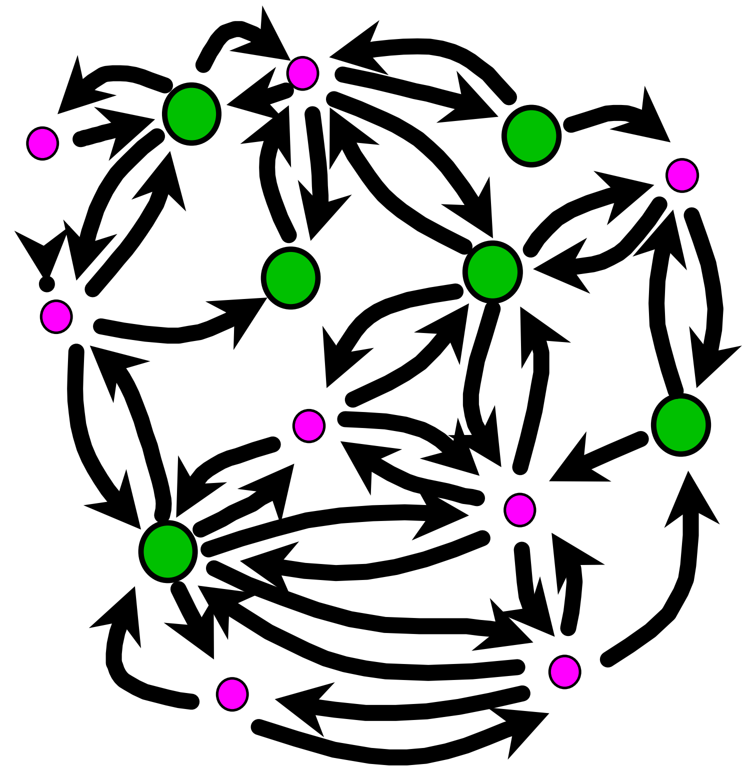
Selecting the coarse-grid points



Select a final **independent set** and perform the heuristics.

Selecting the coarse-grid points

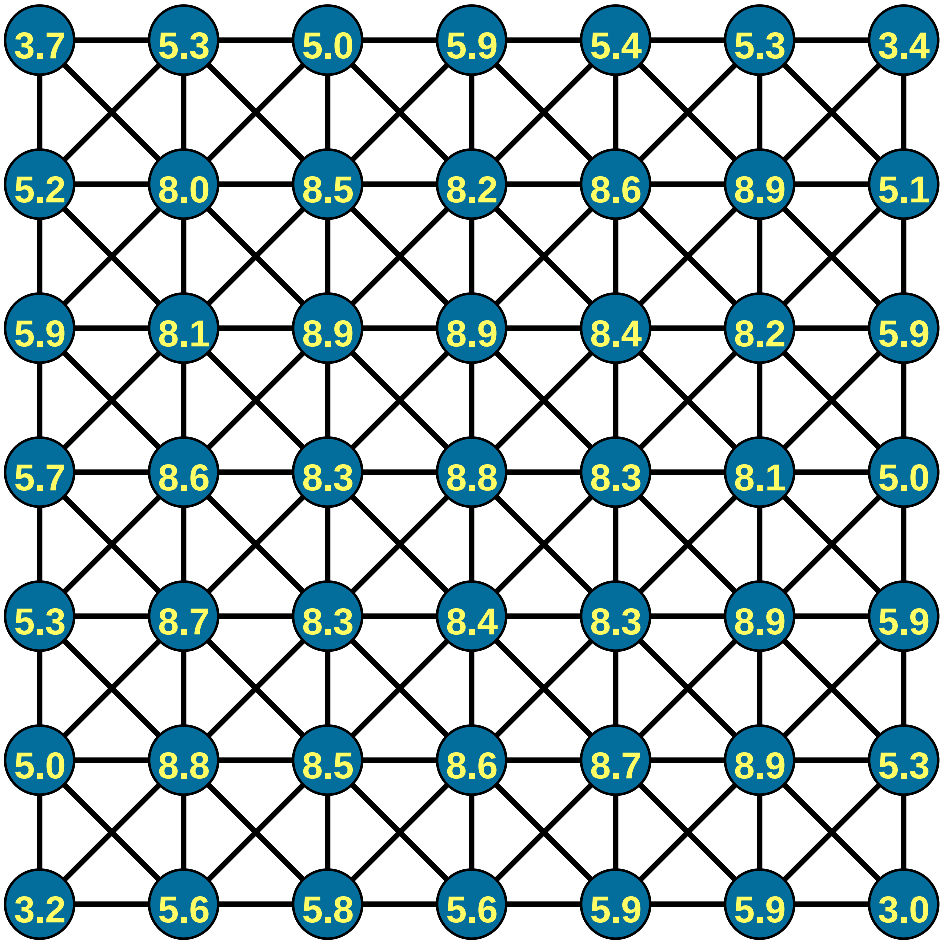




PAMG coarsening:
7 C-points selected

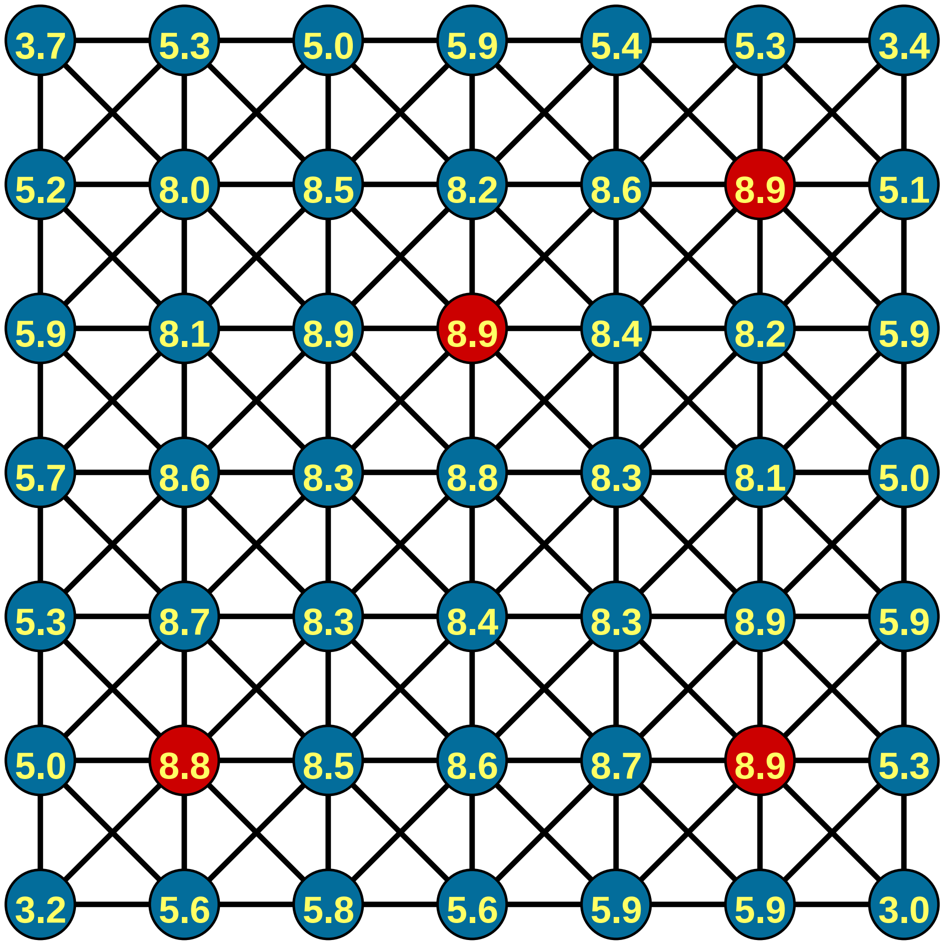
Standard AMG coarsening:
6 C-points selected

ParAMG start



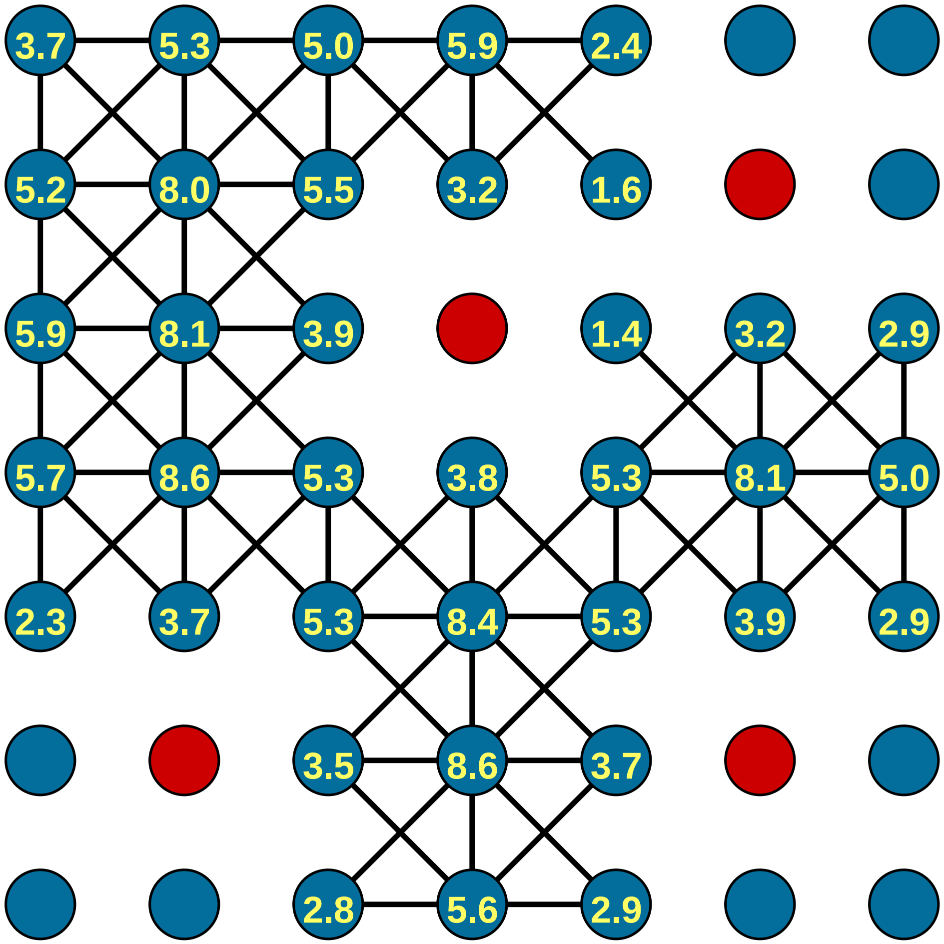
- select C-pts with maximal measure locally
- remove neighbor edges
- update neighbor measures

ParAMG select 1



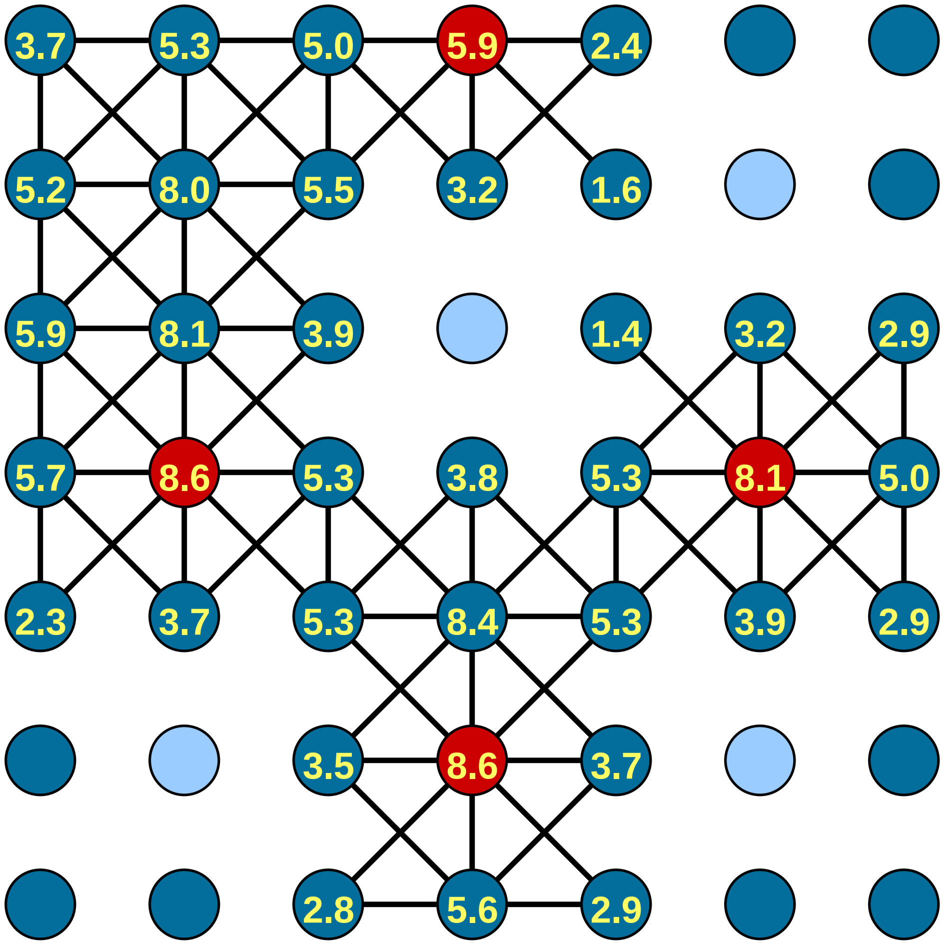
- **select C-pts with maximal measure locally**
- **remove neighbor edges**
- **update neighbor measures**

ParAMG remove and update 1



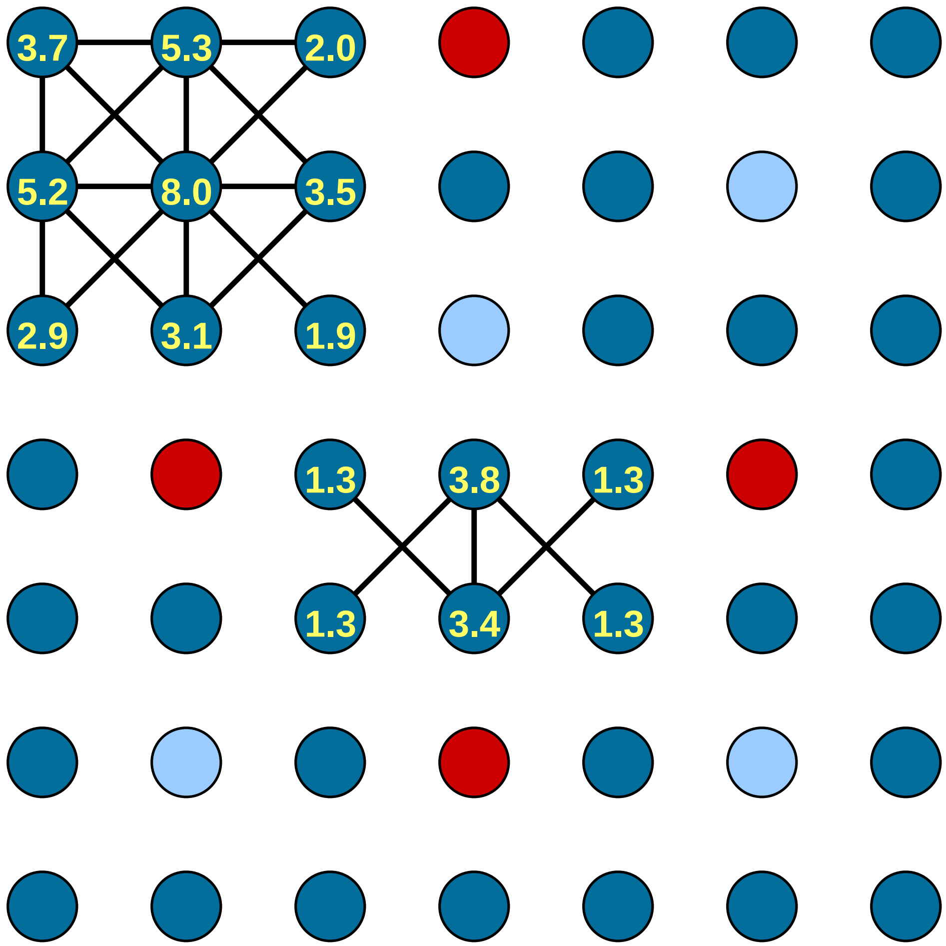
- select C-pts with maximal measure locally
- remove neighbor edges
- update neighbor measures

ParAMG select 2



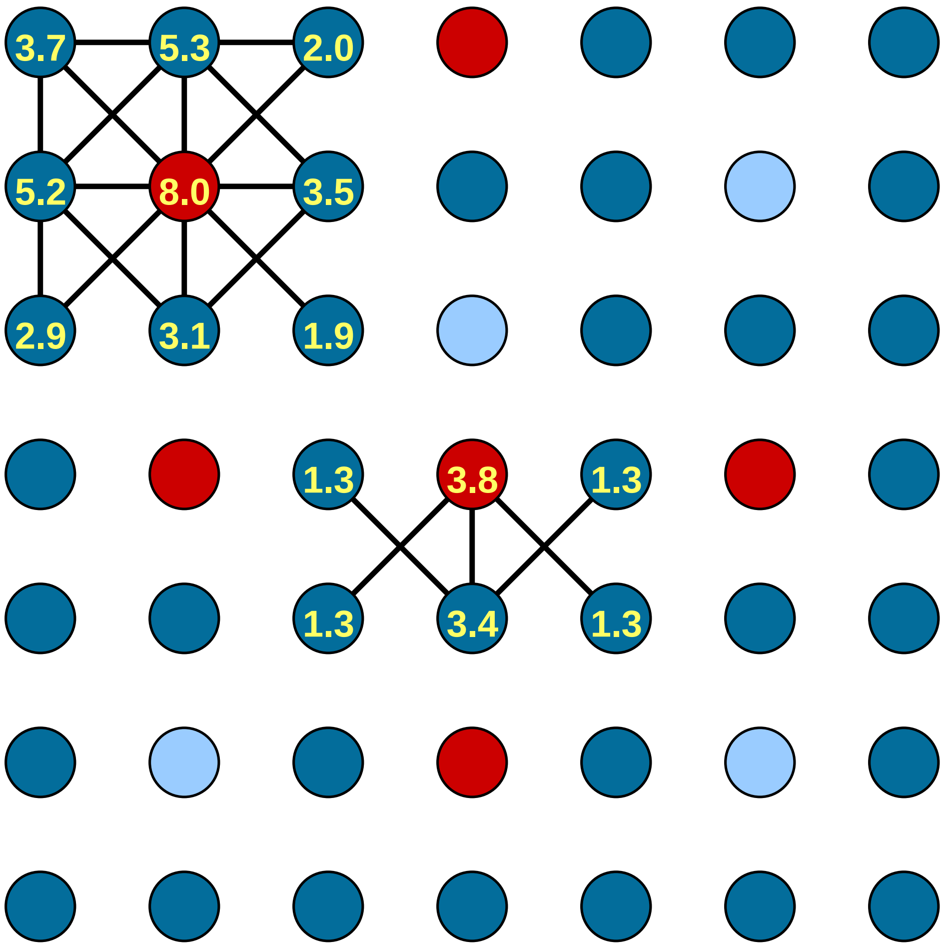
- select C-pts with maximal measure locally
- remove neighbor edges
- update neighbor measures

ParAMG remove and update 2



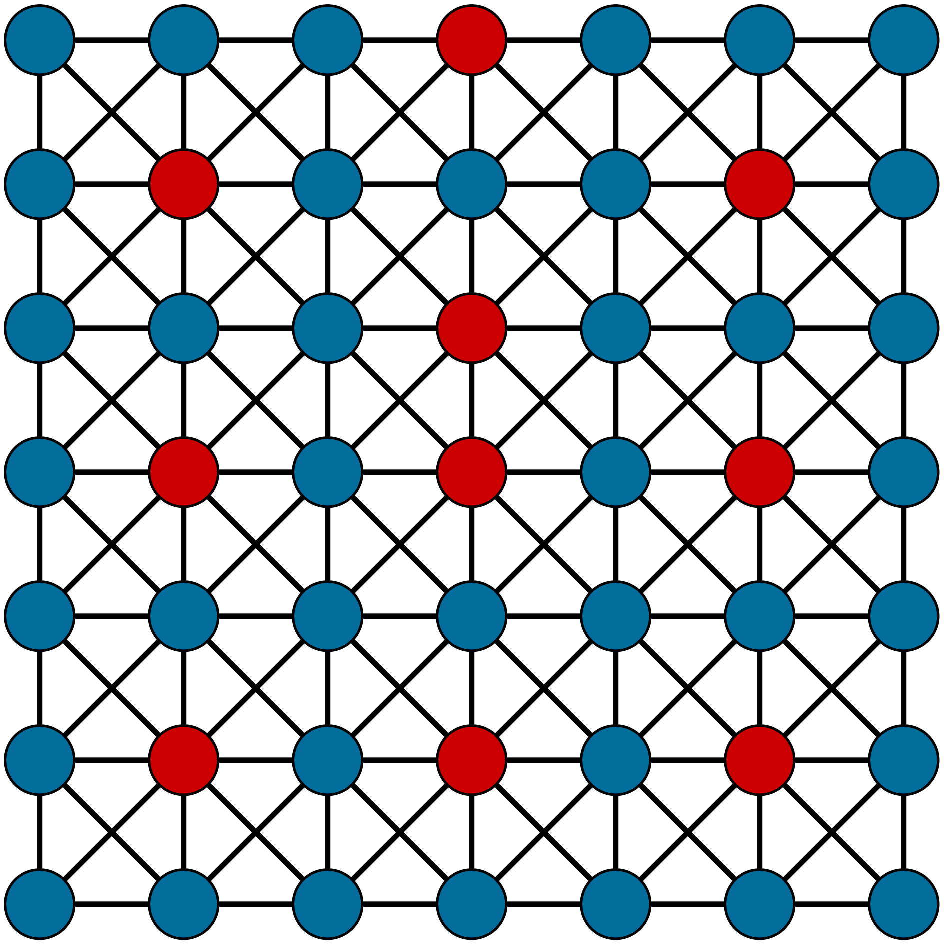
- select C-pts with maximal measure locally
- remove neighbor edges
- update neighbor measures

ParAMG select 3



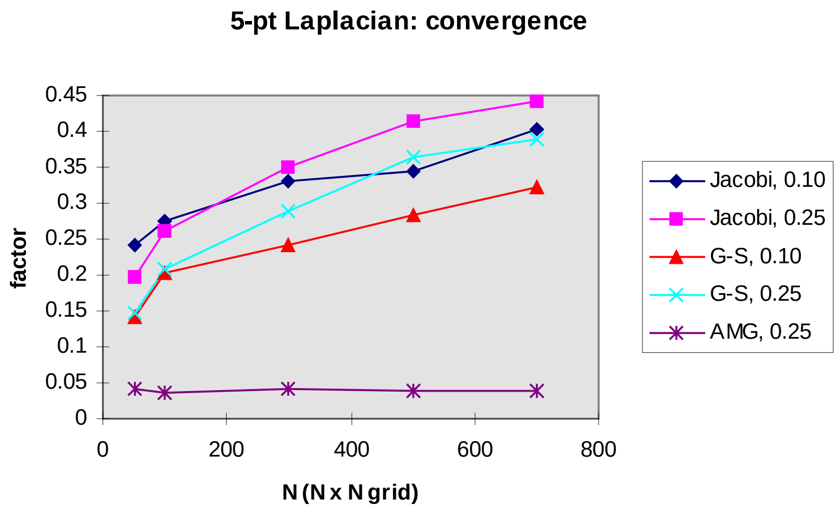
- select C-pts with maximal measure locally
- remove neighbor edges
- update neighbor measures

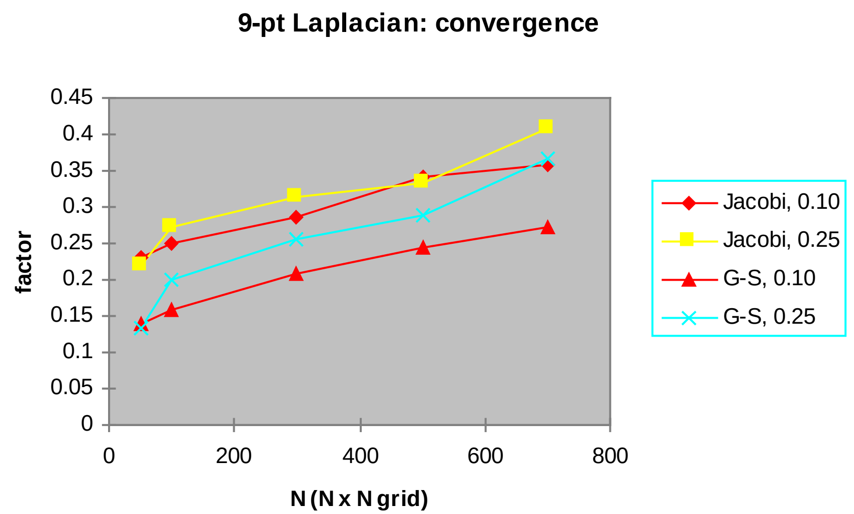
ParAMG final grid

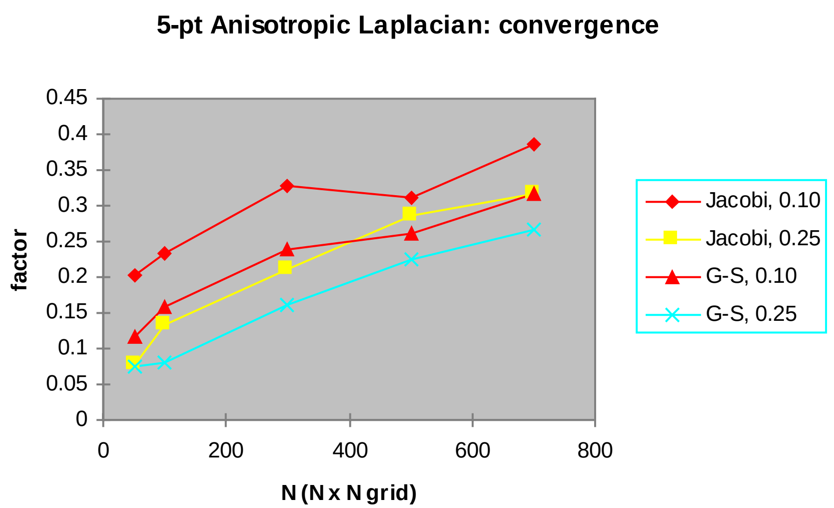


- 11 C-points selected
- Standard AMG selects 9 C-points

ParAMG results: convergence factor

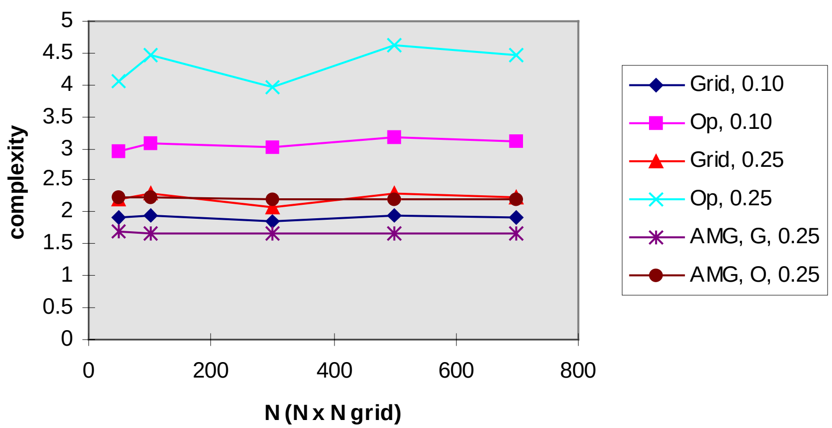




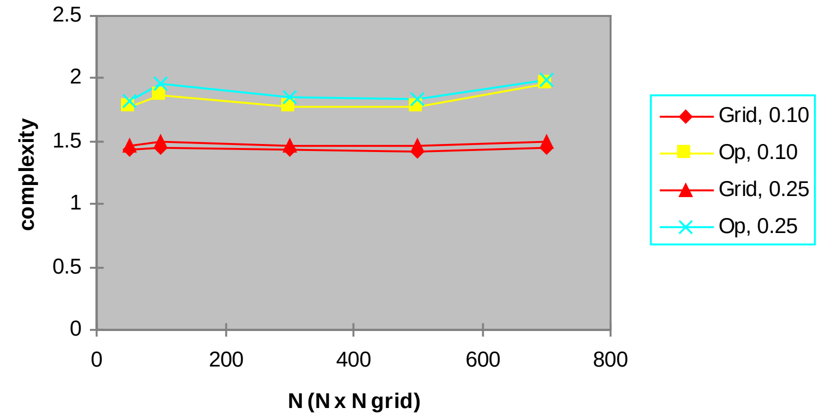


ParAMG results: complexity

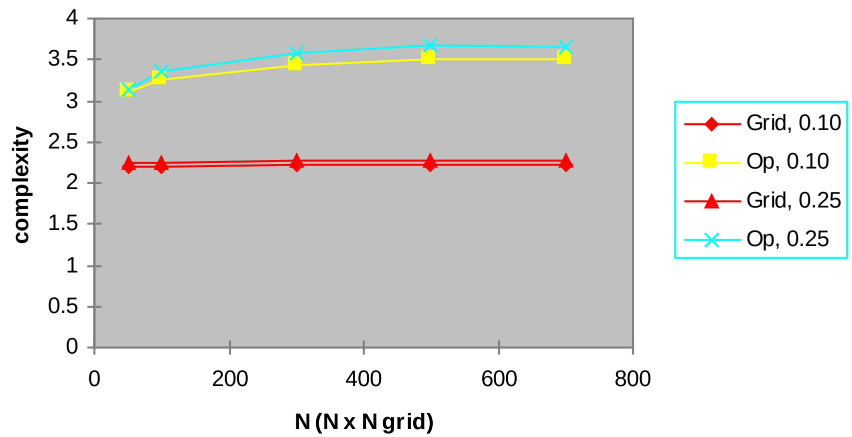
5-pt Laplacian: complexity



9-pt Laplacian: complexity



5-pt Anisotropic Laplacian: complexity



Back to AMG: what else can go wrong? Thin body elasticity!

- **Elasticity, 3-d, thin bodies!**

$$u_{xx} + \frac{1-\nu}{2}(u_{yy} + u_{zz}) + \frac{1+\nu}{2}(v_{xy} + w_{xz}) = f_1$$

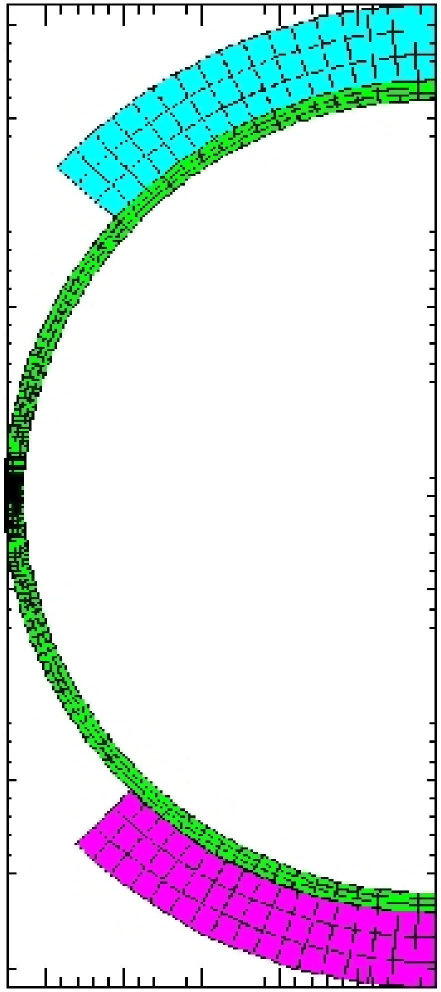
$$v_{yy} + \frac{1-\nu}{2}(v_{xx} + v_{zz}) + \frac{1+\nu}{2}(u_{xy} + w_{yz}) = f_2$$

$$w_{zz} + \frac{1-\nu}{2}(w_{xx} + w_{yy}) + \frac{1+\nu}{2}(u_{xz} + v_{yz}) = f_3$$

- **Slide surfaces, Lagrange multipliers, force balance constraints:**

$$\begin{pmatrix} S & T \\ U & V \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}$$

- ***S*** **is "generally" positive definite,** ***V*** **can be zero,** $U^T \neq T$**.**



We need a more robust characterization of smooth error

- **Example: consider quadrilateral finite elements on a stretched 2D Cartesian grid ($\Delta x \to \infty$)**

$$A = \begin{bmatrix} -1 & -4 & -1 \\ 2 & 8 & 2 \\ -1 & -4 & -1 \end{bmatrix}$$

- **Strong dependence is not apparent here**
- **Iterative weight interpolation will sometimes compensate for mis-identified strong dependence**
- **Elasticity problems are still problematic**

Global measures that relate interpolation accuracy and eigenmodes

- **Fundamental heuristic:** for a two grid algorithm, the interpolation operator must be able to reproduce a mode up to the same accuracy as the size of the associated eigenvalue.
- **That is, one the following should be small**

measure 1:

$$\frac{\left\langle (I-E)e, (I-E)e \right\rangle}{\left\langle Ae, e \right\rangle}$$

measure 2:

$$\frac{\left\langle A(I-E)e, (I-E)e \right\rangle}{\left\langle Ae, Ae \right\rangle}$$

Two-level convergence results

- If *M* is an upper bound for either measure 1 or 2, then the V(1,0) convergence factor is bounded by

$$\alpha \le \left(1 - \frac{1}{M \| A \|} \right)^{1/2}$$

- The V(1,1) convergence factor is the square of above
- A multi-level result for measure 2 can be found in McCormick, SINUM 1985.

AMGe uses f.e. stiffness matrices to characterize smooth error locally

- **For each point *i*, sum the local f.e. stiffness matrices**

$$A_i = \sum_{k \in \eta_i} B_k$$

- **We want the following local measures to be small**

local measure 1:

$$\frac{\left\langle \varepsilon_i \varepsilon_i^T (I - E)\, e, (I - E)\, e \right\rangle}{\left\langle A_i e, e \right\rangle}$$

local measure 2:

$$\frac{\left\langle \varepsilon_i \varepsilon_i^T A_i \varepsilon_i \varepsilon_i^T (I - E)\, e, \,(I - E)\, e \right\rangle}{\left\langle A_i e, A_i e \right\rangle}$$

We use the local measures to define interpolation

- Given a coarse grid, we define interpolation to be the *arg min* of (for measure 1)

$$\min_E \max_e \frac{\left\langle \varepsilon_i \varepsilon_i^T (I - E)\, e, (I - E)\, e \right\rangle}{\left\langle A_i e, e \right\rangle}$$

- Similarly for measure 2

Using local measures to define interp. is equivalent to fitting local eigenmodes

- **Assume the eigen-decomposition:**

$$A_i V = \Lambda V; \quad \Lambda = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_+ \end{bmatrix}; \quad V = \begin{bmatrix} V_0 & V_+ \end{bmatrix};$$

- **Finding the *arg min* of measures 1 and 2 is equivalent to solving the following constrained least-squares problems with *p=1/2* and *p=1***

$$\min_w \left\| \Lambda_+^{-p} V_+^T \begin{pmatrix} 1 \\ 0 \\ -w \end{pmatrix} \right\|^2, \qquad \text{s.t.} \quad V_0^T \begin{pmatrix} 1 \\ 0 \\ -w \end{pmatrix} = 0$$

Computing interpolation in practice

- **For measure 1, partition local matrix by *F* and *C*-pts:**

$$A_i = \begin{bmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{bmatrix}$$

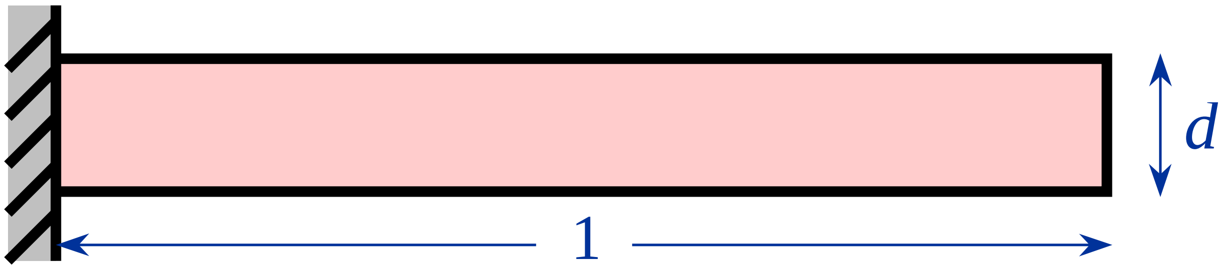
- **Then interpolation to point *i* is defined by**

$$w^T = \varepsilon_i^T W, \qquad \text{s.t. } A_{ff} W = -A_{fc}$$

- **Measure 1 seems to produce better interpolation in practice than does measure 2**

Preliminary results for new AMGe prolongation are promising

- 2D plane-stress cantilever beam, fixed on one end.



- **Convergence factors:** finest grid elements are $h \times h$; grids are coarsened geometrically.

d	h	AMG prolongation	AMGe prolongation
1	1/32	0.60	0.20
1/4	1/8	0.95	0.25
1/8	1/16	0.90	0.26
1/16	1/64	0.92	0.26

Bounds on local measure 1 (and 2) yield global convergence bounds

$$
\begin{aligned}
\left\langle (I-E)e, (I-E)e \right\rangle &= \sum_{i \in F} \left\langle \varepsilon_i \varepsilon_i^T (I-E)e, (I-E)e \right\rangle \\
&\le \sum_{i \in F} M_i \left\langle A_i e, e \right\rangle \\
&= \sum_k \left\langle B_k e, e \right\rangle \sum_{i \in N_k \cap F} M_i \\
&\le M \left\langle Ae, e \right\rangle
\end{aligned}
$$

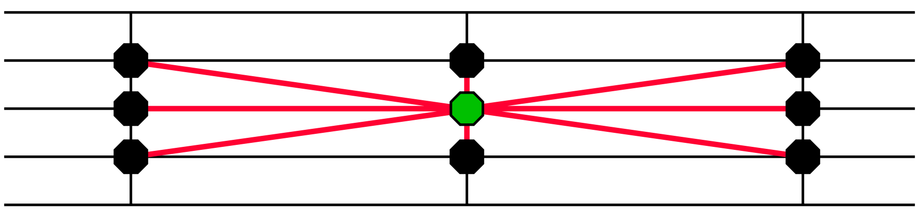
- **We would like to use this bound to assess the quality of interpolation and coarse-grid choice**

$$
\alpha \le \left(1 - \frac{1}{M \| A \|} \right)^{1/2}
$$

Using the local measure to define "strong dependence"

- **Consider interpolating to point *i* independently from each of its neighbors:**

For each point *k* connected to point *i* , compute value of M_i assuming *k* is the only C-point.



$$A = \begin{bmatrix} -1 & -4 & -1 \\ 2 & 8 & 2 \\ -1 & -4 & -1 \end{bmatrix}$$

M_i values for quadrilateral example (1000:1)

3×10^6	9.7	3×10^6
3×10^6		3×10^6
3×10^6	9.7	3×10^6

Target Problems:

- Adaptive, Lagrangian-Eulerian 3d code :
  - Radiation-Hydro, manufacturing, gross deformation
  - slide surfaces & thin bodies
- We have found that AMG-preconditioned GMRES can be used effectively on several typical problems
- We hope to test AMGe ideas out on several test problems soon
  - now able to get stiffness matrices
  - will use a modified AMG coarsening heuristic to coarsen appropriately near boundaries
  - look first at a problem without slide surfaces

Conclusions

- *AMG has been shown to be a robust, efficient solver on a wide variety of problems of real-world interest.*
- **AMG can be parallelized using a modified Luby-Jones-Plassman parallel MIS algorithm to develop the parallel coarse-grid selection.**
- **Tests show that the parallel algorithm will produce grids that give acceptable (not optimal) convergence factors. Future work includes modifying the coarsening algorithm to improve performance and ensure scalability.**
- **Finite element stiffness matrices and local measures can be used to better characterize smooth error, select coarse grids, and construct interpolation.**

Acknowledgement

- This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.